UNIVERSIDAD CARLOS III DE MADRID

working papers

UC3M Working Papers
Statistics and Econometrics
17-08
ISSN 2387-0303
Mayo 2017

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

ESTIMATION OF A DYNAMIC MULTI-LEVEL FACTOR MODEL WITH POSSIBLE LONG-RANGE DEPENDENCE

Yunus Emre Ergemen[a], Carlos Vladimir Rodríguez-Caballero[b]

Abstract

A dynamic multilevel factor model with possible stochastic time trends is proposed. In the model, long-range dependence and short memory dynamics are allowed in global and regional common factors as well as model innovations. Estimation of global and regional common factors is performed on the prewhitened series, for which the prewhitening parameter is estimated semiparametrically from the cross-sectional and regional average of the observable series. Employing canonical correlation analysis and a sequential least-squares algorithm on the prewhitened series, the resulting multilevel factor estimates have a centered asymptotic normal distribution. Selection of the number of global and regional factors is also discussed. Estimates are found to have good small-sample performance via Monte Carlo simulations. The method is then applied to the Nord Pool electricity market for the analysis of price comovements among different regions within the power grid. The global factor is identified to be the system price, and fractional cointegration relationships are found between regional prices and the system price.

Keywords: Multi-level factor; long-range dependence; short memory; fractional cointegration; Nord Pool power market.

[a] CREATES and Aarhus University, Department of Economics and Business Economics.
E-mail: ergemen@econ.au.dk. Fuglesangs Allé 4, 8210 Aarhus V, Denmark.
b Universidad Carlos III de Madrid, Department of Statistics and CREATES.
E-mail: carlosvladimir.rodriguez@uc3m.es. Calle Madrid, 126, 28903 Getafe, Madrid, Spain.

Acknowledgements: We are grateful to Niels Haldrup, Carlos Velasco, Jesús Gonzalo, Esther Ruiz, Javier Hualde, and the participants in CREATES Seminars in 2015 and 2016, the V It Workshop in Time Series Econometrics in Zaragoza, the International Association for Applied Econometrics 2016 Annual Conference, and the CREATES Anniversary Meeting for their constructive comments. We also acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

Estimation of a Dynamic Multi-Level Factor Model with Possible Long-Range Dependence*

Yunus Emre Ergemen†
Carlos Vladimir Rodríguez-Caballero‡

March 27, 2017

Abstract

A dynamic multilevel factor model with possible stochastic time trends is proposed. In the model, long-range dependence and short memory dynamics are allowed in global and regional common factors as well as model innovations. Estimation of global and regional common factors is performed on the prewhitened series, for which the prewhitening parameter is estimated semiparametrically from the cross-sectional and regional average of the observable series. Employing canonical correlation analysis and a sequential least-squares algorithm on the prewhitened series, the resulting multilevel factor estimates have a centered asymptotic normal distribution. Selection of the number of global and regional factors is also discussed. Estimates are found to have good small-sample performance via Monte Carlo simulations. The method is then applied to the Nord Pool electricity market for the analysis of price comovements among different regions within the power grid. The global factor is identified to be the system price, and fractional cointegration relationships are found between regional prices and the system price.

Keywords: Multi-level factor; long-range dependence; short memory; fractional cointegration; Nord Pool power market.
JEL Classification: C12, C22.

*Acknowledgments. We are grateful to Niels Haldrup, Carlos Velasco, Jesús Gonzalo, Esther Ruiz, Javier Hualde, and the participants in CREATES Seminars in 2015 and 2016, the VI_t Workshop in Time Series Econometrics, the International Association for Applied Econometrics 2016 Annual Conference, and the CREATES Anniversary Meeting for their constructive comments. We also acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†Corresponding author. CREATES and Aarhus University, Department of Economics and Business Economics. E-mail: ergemen@econ.au.dk. Fuglesangs Allé 4, 8210 Aarhus V, Denmark.

‡Universidad Carlos III de Madrid, Department of Statistics and CREATES. E-mail: carlosvladimir.rodriguez@uc3m.es. Calle Madrid, 126, 28903 Getafe, Madrid, Spain.

1 Introduction

Dynamic factor models are extensively used as a dimension reduction tool in the analysis of large economic data sets. For estimation and inference theory under different setups, see e.g. Stock and Watson (2002), Bai and Ng (2002), Bai (2003), Bai and Ng (2004), and Bai and Ng (2008). While there is a vast literature for estimation of the common factors and the number of common factors when both cross-section and time-series dimensions are large, most available methodologies rely only on the existence of pervasive factors.

More recently there has been some interest in decomposing common factor structures into different levels. The intuition behind a multi-level factor structure is intrinsically related to the well-known Tobler's first law of geography, "*everything is related to everything else, but near things are more related than distant things*", which is the foundation of many ideas embodied in spatial statistics. In this sense, a multi-level factor structure is based on a decomposition of the factor space in global and regional components. While the global factors capture common movements between all regions, the regional components capture only those that are unique to specific regions. Standard (uni-level) factor analysis is quite limited when there is an interest also in regional rather than just global dynamics.

In this paper, we use a special restricted factor model which mainly differs from the conventional setups in that the model implies lots of zero restrictions in the associated loading matrix. Wang (2010) and Choi et al. (2016) consider this kind of multi-level factor structure under a stationary $I(0)$ setup for which identification is discussed and inference theory is developed. Another related approach of a multi-level factor model is given by imposing a hierarchical structure, see e.g. Moench et al. (2013) and Diebold et al. (2008). The hierarchical approach divides each block of data into some sub-blocks to characterize the within and between-block variations and arrive at the hierarchical (multi-level) model which makes it impossible to separately identify global and regional factors in contrast to the methodology proposed in this paper, see also Wang (2010), Breitung and Eickmeier (2016), and Choi et al. (2016).

We consider a dynamic multi-level factor model that allows for both pervasive (or global) and nonpervasive (e.g. regional) common factors. These common factors are allowed to exhibit long-range dependence and short memory without $I(0)$ stationarity or $I(1)$ nonstationarity restrictions as traditionally imposed in the literature. Model innovations are also allowed to exhibit long-range dependence and short memory properties. This way, not only does the model have greater flexibility but also a priori unit-root or stationarity testing is not needed. Furthermore, many economic and financial indicators, such as production, price and rates series, may exhibit fractional long-range dependence; see e.g. Gil-Alaña and Robinson (1997), Michelacci and Zaffaroni (2000), Bollerslev et al. (2013) and Pesaran and Chudik (2014), and our model is well suited for, but not limited to, the study of such indicators.

In the estimation, we first use the exact local Whittle method of Shimotsu (2010), which remains agnostic about the underlying short-memory dynamics, to estimate the prewhitening parameter with which we difference the observable series to later apply a sequential procedure to estimate the global and regional factors. The estimation method is similar in spirit to that of Breitung and Eickmeier (2016) and Choi et al. (2016). We establish the asymptotic behavior of the multilevel factor structure and prewhitening parameter estimates. We also discuss selection of the number of global and regional common factors based on information criteria proposed by Bai and Ng (2002) and Alessi et al. (2010) using tools from set theory.

Monte Carlo simulations show that the methodology works well even in relatively small panels. We then apply the methodology to study the complex dynamics of the Nord Pool power market in a large panel of hourly observations, for which the global and regional factors drive the commonality overall and among bidding areas in the Nord Pool power market, respectively. We find that the global factor can be interpreted as the system price and that there are fractional cointegrating relationships between regional prices and the system price.

Next section introduces the model along with the model assumptions and contains the estimation strategy along with the corresponding inferential theory. Section 3 discusses the selection of the number of global and regional factors. Section 4 presents a finite sample study based on Monte Carlo simulations. Section 5 provides an empirical application to the Nord Pool energy market, and finally Section 6 concludes the paper.

Throughout the paper, $\|A\| = (trace(A'A))^{1/2}$ for a matrix A, $\mathbf{x}_n = O_p(\mathbf{y}_n)$ states that the vector of random variables, $\mathbf{x}_n$, is at most of order $\mathbf{y}_n$ in probability, and $\mathbf{x}_n = o_p(\mathbf{y}_n)$ is of smaller order in probability than $\mathbf{y}_n$, $\to_p$ denotes convergence in probability, and $\to_d$ denotes convergence in distribution, and $(N, T)_j$ denotes the joint cross-section and time-series asymptotics. All mathematical proofs are collected in an appendix at the end of the paper.

2 The two-level factor model with possible long-range dependence

2.1 The model

We consider a two-level factor model in that the unobserved common shocks are classified into two types: the first factor type is the global or common factor, which is a pervasive top-level factor that affects all economic sectors or regions; the second factor type is the regional or sector-specific factor, which is the nonpervasive sub-level factor and affects only a particular sector or region. Many macroeconomic applications label such factors as global and regional factors, and in this paper we use the terms global/pervasive/common factor and regional/nonpervasive /sector-specific factor interchangeably.

Let $y_{r,it}$ be the observation on the region r, the cross-section unit i at time t for $r = 1, \ldots, R$; $i = 1, \ldots, N_r$; $t = 1, \ldots, T$ that is generated by

$$y_{r,it} = \mu'_{r,i} G_t + \lambda'_{r,i} F_{r,t} + \epsilon_{r,it}. \tag{1}$$

In the model, the total number of observations across all regions is $N = N_1 + N_2 + \cdots + N_R$. We take the number of regions R to be fixed since this is generally, if not always, enough from a practical point of view and it makes the asymptotic analysis much more tractable. The $\mathbf{r}_G \times 1$ vector $G_t = (g_{1,t}, \ldots, g_{\mathbf{r}_G,t})'$ contains the $\mathbf{r}_G$ unobservable global factors and the $\mathbf{r}_{F_r} \times 1$ vector $F_{r,t}$ consists of the $\mathbf{r}_{F_r}$ unobservable regional factors in region r. Naturally, the number of regional factors can be different in each region. $\mu_{r,i}$ and $\lambda_{r,i}$ are $\mathbf{r}_G$- and $\mathbf{r}_{F_r}$-dimensional factor loadings showing how each unit i in region r is affected by G_t and $F_{r,t}$, respectively.

The intuition behind a multilevel factor model is that each process $y_{r,it}$ is the sum of a *global common component*, a *regional common component*, and an idiosyncratic component. Common components of region r are driven by the respective $\mathbf{r}_G$ and $\mathbf{r}_F$ vectors of common factors (global and regional), which are possibly loaded differently. For practical purposes, there may be an interest in measuring certain comovements between countries employing multilevel factors. In that case, the global component would capture common movements in all groups of countries, and the regional component would capture common movements with the country's neighbors whereas the specific country component would capture movements that are unique to that specific country. Comovements between countries as captured by these multilevel factors can then be used to measure the connectivity of the countries analyzed. For instance, if the regional component of a specific country weighs more than the global component, the country would seem to be more connected with its neighbors than with all the countries as a whole.

In (1),

$$
\begin{aligned}
G_t &= \Delta_t^{-\delta_0} w_t, \\
F_{r,t} &= \Delta_t^{-\vartheta_{r,0}} v_{r,t}, \text{ and} \\
\epsilon_{r,it} &= \Delta_t^{-d_{r,i0}} u_{r,it},
\end{aligned}
$$

where w_t, $v_{r,t}$, and, $u_{r,it}$ are stationary $I(0)$ processes with spectral densities $f_w(\omega) \sim G_w$, $f_{rv}(\omega) \sim G_{rv}$ and $f_{riu}(\omega) \sim G_{riu}$ when $\omega \sim 0$ for $r = 1, \ldots, R$ and $i = 1, \ldots, N_r$.

With $\Delta = 1 - L$, and L such that $L^s x_t = x_{t-s}$, $\Delta^{-\zeta}$ has the expansion

$$
\Delta^{-\zeta} = \sum_{j=0}^{\infty} \pi_j(-\zeta) L^j, \quad \text{where } \pi_j(-\zeta) = \frac{\Gamma(j+\zeta)}{\Gamma(j+1)\Gamma(\zeta)},
$$

for $\zeta > 0$ with $\Gamma(\tau) = \infty$ for $\tau = 0, -1, \ldots$, and $\Gamma(0)/\Gamma(0) = 1$. $\Delta_t^{-\zeta}$ truncates this filter as $\Delta_t^{-\zeta} = \sum_{j=0}^{t} \pi_j(-\zeta) L^j$, and this truncation allows for the study of both stationary $(\zeta < 1/2)$ and the nonstationary $(\zeta \geq 1/2)$ cases, unlike the untruncated filter $\Delta^{-\zeta}$ that does not converge when $\zeta \geq 1/2$, see Davidson and Hashimzade (2009).

Multilevel factor models, such as (1), can prove difficult to identify, see also Wang (2010). As one strategy, we impose a block of zero restrictions on the matrix of factor loadings. This way, the system of all R regions can be represented as

$$
\begin{pmatrix} y_{1,\cdot t} \\ \vdots \\ y_{R,\cdot t} \end{pmatrix} = \begin{pmatrix} \mu_1 & \Lambda_1 & 0 & \cdots & 0 \\ \mu_2 & 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_R & 0 & 0 & \cdots & \Lambda_R \end{pmatrix} \begin{pmatrix} G_t \\ F_{1,t} \\ F_{2,t} \\ \vdots \\ F_{R,t} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,\cdot t} \\ \vdots \\ \epsilon_{R,\cdot t} \end{pmatrix},
$$

$$
y_t = \Lambda^* F_t^* + \epsilon_t, \tag{2}
$$

where $F_t^* = \left(G_t', F_{1,t}', \ldots, F_{R,t}'\right)'$ and $\Lambda^* = [\mu_r, \Lambda_r]$ with Λ_r block diagonal. The full system could be written in matrix form as

$$Y = F^* \, \Lambda^{*'} + E,$$

where the dimension of Y, F_t^*, and Λ^* being $T \times N$, $T \times (\mathbf{r}_{F_1} + \cdots + \mathbf{r}_{F_R} + \mathbf{r}_G)$, and $N \times (\mathbf{r}_{F_1} + \cdots + \mathbf{r}_{F_R} + \mathbf{r}_G)$, respectively.

We introduce the following conditions to study to (1) letting $\mathcal{M}$ denote a generic positive constant.

Assumption A. Long-range dependence and short-memory dynamics:

A_1 $f_w(\omega) \sim G_{w0} \in (0, \infty)$ and also for some $\beta \in (0, 2]$, $f_w(\omega) = G_{w0}(1 + O(\omega^\beta))$ as $\omega \to 0+$. Furthermore, $f_w(\omega)$ is bounded for $\omega \in [0, \pi]$. These conditions also hold for $f_{rv}(\omega)$ with G_{rv} and $f_{riu}(\omega)$ with G_{riu}.

A_2 $w_t = A(L)w_t = \sum_{k=0}^{\infty} a_k w_{t-k}$ with $\sum_{k=0}^{\infty} a_k^2 < \infty$, where $E(w_t|\mathcal{F}_{t-1}) = 0$, $E(w_t^2|\mathcal{F}_{t-1}) = 1$, $E(w_t^3|\mathcal{F}_{t-1}) = \mu_3$, $E(w_t^4|\mathcal{F}_{t-1}) = \mu_4$, with finite constants μ_3, μ_4 almost surely, $t = 0, \pm 1, \ldots$, in which $\mathcal{F}_t$ is the $\sigma-$field generated by w_s, $s \le t$, and there exists a random variable w such that $Ew^2 < \infty$ and for all $\eta > 0$ and some $K > 0$, $Pr(|w_t| > \eta) \le KPr(|w| > \eta)$. Similarly, for $v_{rt} = B_r(L)z_{rt} = \sum_{k=0}^{\infty} b_{rk} z_{r,t-k}$ with $\sum_{k=0}^{\infty} b_{rk}^2 < \infty$ and for $u_{rit} = C_{ri}(L)\epsilon_{rit} = \sum_{k=0}^{\infty} c_{rik}\epsilon_{ri,t-k}$ with $\sum_{k=0}^{\infty} c_{rik}^2 < \infty$ for each r and i, the same conditions hold.

A_3 In a neighborhood $(0, \kappa)$ of the origin, $A(e^{i\omega})$, $B_r(e^{i\omega})$ and $C_{ri}(e^{i\omega})$, with i in the exponent s.t. $i^2 = -1$, are differentiable, $(d/d\omega)A(e^{i\omega}) = O(\omega^{-1})$, $(d/d\omega)B_r(e^{i\omega}) = O(\omega^{-1})$ and $(d/d\omega)C_{ri}(e^{i\omega}) = O(\omega^{-1})$ as $\omega \to 0+$.

A_4 Denoting m a bandwidth parameter, as $T \to \infty$, $m^{-1} + m^{1+2\beta}(\log m)^2 T^{-2\beta} + m^{-\gamma}\log T \to 0$ for any $\gamma > 0$.

A_5 $\delta_0 \in \mathcal{G} = [\underline{\delta}, \bar{\delta}]$ and $-1/2 < \underline{\delta} < \bar{\delta} \le 7/4$. Also, for $r = 1, \ldots, R$, with R fixed, $\vartheta_{r0} \in \mathcal{V}_r = [\underline{\vartheta}_r, \bar{\vartheta}_r]$ and $-1/2 < \underline{\vartheta}_r < \bar{\vartheta}_r \le 7/4$, and for $i = 1, \ldots, N_r$, $d_{ri0} \in \mathcal{D}_{ri} = [\underline{d}_{ri}, \bar{d}_{ri}]$ and $-1/2 < \underline{d}_{ri} < \bar{d}_{ri} \le 7/4$.

Assumption B. Factors:
Denoting $G_t^0 = \Delta_t^{\delta_0} G_t$ and $F_{r,t}^0 = \Delta_t^{\vartheta_{r,0}} F_{r,t}$, define $H_t = \left[G_t^{0'}, F_{r,t}^{0'}\right]'$. For a fixed

$\mathbf{r}$, assume that $T^{-1}\sum_{t=1}^{T} H_t H_t^{'} \xrightarrow{p} \sum_H$ for some positive-definite matrix $\sum_H$, as $T \to \infty$ with rank $\mathbf{r}_G + \mathbf{r}_{F_1} + \cdots + \mathbf{r}_{F_R}$.

Let $\lambda_{r,i}^0$ and $\mu_{r,i}^0$ denote the true regional and global factor loadings, respectively.

Assumption C. Factor loadings:

C_1 $\lambda_{r,i}^0$ is either deterministic such that $||\lambda_{r,i}^0|| \leq \mathcal{M} < \infty$, or it is stochastic such that $E||\lambda_{r,i}^0||^4 \leq \mathcal{M} < \infty$. In the latter case, $N_r^{-1}\Lambda_r^{0'}\Lambda_r^0 \xrightarrow{p} \Sigma_{\Lambda_r} > 0$ for an $\mathbf{r}_F \times \mathbf{r}_F$ non-random matrix Σ_{Λ_r}, as $N_r \to \infty$ for all $r = 1, \ldots, R$.

C_2 $\mu_{r,i}^0$ is either deterministic such that $||\mu_{r,i}^0|| \leq \mathcal{M}$, or it is stochastic such that $E||\mu_{r,i}^0||^4 \leq \mathcal{M} < \infty$ with $N_r^{-1}\mu_r^{0'}\mu_r^0 \xrightarrow{p} \Sigma_{\mu_r} > 0$ for an $\mathbf{r}_G \times \mathbf{r}_G$ non-random matrix Σ_{μ_r}, as $N_r \to \infty$ for all $r = 1, \ldots, R$.

C_3 Rank $([\mu_\mathbf{r}\ \Lambda_\mathbf{r}]) = \mathbf{r}_G + \mathbf{r}_{F_r}$.

Assumption D. Processes $\{u_{r,it}\}$, $\{v_{r,t}\}$, $\{w_t\}$, $\{\lambda_{r,i}\}$, and $\{\mu_{r,i}\}$ are mutually independent groups.

Assumption E. Identification:

E_1 $F_r^{'}F_r/T = I_{\mathbf{r}_{F_r}}$ and $\Lambda_r^{'}\Lambda_r$ diagonal (Within region identification).

E_2 $G^{'}G/T = I_{\mathbf{r}_G}$ and $\mu^{'}\mu$ diagonal (Between region identification).

E_3 Factors have zero mean, and $\sum_{t=1}^{T} G_t F_{r,t}^{'} = 0$ for $r = 1, \ldots, R$.

Assumption A imposes the restrictions used by Shimotsu (2010) for exact local Whittle estimation incorporating possible unknown mean and polynomial trend while being agnostic about the underlying short-run dynamics. The allowed range of memory values greatly relaxes the $I(0) - I(1)$ restrictions vastly imposed in the factor literature. The model in (1) simultaneously admits combination of persistence levels in factors as well as in the idiosyncratic terms. Hence, Assumption A permits extensive fractional cointegrating restrictions on the model and can be useful in understanding the behavior of co-persistent indicators involved in the dynamics of a complex system although a cointegrating relationship is not a priori imposed. As a result, most factor models in the literature, such as those proposed by Stock and Watson (2002), Bai and Ng (2002, 2004) and Wang (2010) are readily nested under (1). For further discussion on the conditions in Assumption A, readers are referred to Shimotsu (2010).

Assumption B is a standard condition that states that the fully whitened factors have a positive-definite variance-covariance matrix as $T \to \infty$. The rank condition in Assumption B implies that different factors are not perfectly correlated.

Assumption C_1 ensures that the global factor G_{st} has nontrivial contribution to the variance of y_t, $s = 1, \ldots, r$ while assumption C_2 ensures that each regional factor $F_{r,jt}$ has a nontrivial contribution to the variance of $y_{r,t}$, $j = 1, \ldots, r_F$. The latter means that G_t pervades all variables whereas the regional factor $F_{r,jt}$ pervades only within region r. The rank condition in Assumption C_3 guarantees enough heterogeneity among individual variables within region r when responding to both factors. Such a rank condition is useful in separating identification of global and regional factors. For estimation purposes, we assume the number of factors $\mathbf{r}_G$ and $\mathbf{r}_F$ to be known and fixed at this step. Formal tests or information criteria for such number of factors are, to our knowledge, not yet available even discarding long-range dependence in the factors. Notwithstanding, the number of factors does not affect asymptotic results for the common component, see Bai (2003).

Assumption D implies that unobservable factors, factor loadings, and error components are assumed to be independent of each other. Nevertheless, regional factors from different regions can still be correlated.

We impose the three conditions in Assumption E to identify the factors. Assumptions E_1 and E_2 are standard in factor analysis and allow the model to be uniquely identified under such normalizations. Assumption E_3 rules out any possibility of correlation between the global and regional factors, which is the same as saying that the global factors do not contain information about regional factors and vice versa. This assumption enables us to separately identify regional factors and global factors. Readers are referred to Wang (2010) for further discussion on the restrictions involved in both assumptions.

Following Breitung and Eickmeier (2016) and Proposition 1 in Wang (2010), the factor loadings for the model 2 are identified up to a linear transformation of the loading matrix that preserves the same zero restrictions of the model given by $\Lambda^* Q$ with

$$Q = \begin{pmatrix} Q_{00} & 0 & 0 & \cdots & 0 \\ Q_{10} & Q_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_{R0} & 0 & 0 & \cdots & Q_{RR} \end{pmatrix}, \tag{3}$$

where orthonormal global and regional factors within each of the $R + 1$ blocks are given by $Q_{00} = \left(T^{-1}\sum_{t=1}^{T} G_t G_t^{'}\right)^{-1/2}$ and $Q_{rr} = \left(T^{-1}\sum_{t=1}^{T} F_{r,t} F_{r,t}^{'}\right)^{-1/2}$ for all r. Matrix 3 imposes that the R blocks of regional factors are uncorrelated with the blocks of global factors.

2.2 Estimation

In the estimation, we first obtain the memory of the cross-sectionally and regionally averaged series to prewhiten the series, following a similar reasoning to first differencing used by Bai and Ng (2004). Then we adopt the estimation procedure proposed by Breitung and Eickmeier (2016) to estimate the prewhitened global and regional common factors. This procedure is also discussed for the estimation of $I(0)$ multilevel factors by Choi et al. (2016). A similar procedure related to the sequential PC approach is also proposed by Wang (2010) but that of Breitung and Eickmeier (2016) proves to be computationally simpler. Once the unobservable factors in both levels are estimated, they are integrated back by the initial prewhitening order and memory parameters of the factor estimates are obtained by the Extended Local Whittle (ELW) method, proposed by Abadir et al. (2007), which consistently estimates the memory parameter allowing for a wide range of values. We discuss these steps in detail as follows.

To obtain an estimate of the prewhitening parameter, both parametric, see e.g. Ergemen and Velasco (2017), and semiparametric, see e.g. Shimotsu (2010), methods can be employed. The advantage of semiparametric methods is that they can handle underlying short-memory dynamics in an agnostic way and are robust to possible misspecification, which proves to be important under our setup since both the multilevel factor structure and idiosyncratic errors are unobservable. With this in mind, we estimate the prewhitening parameter from the cross-sectionally and regionally averaged series employing the exact local Whittle estimation due to Shimotsu (2010). To give further details, the cross-sectional and regional average of (1),

$$\bar{\bar{y}} := \frac{1}{N}\frac{1}{R}\sum_{r=1}^{R}\sum_{i=1}^{N_r} y_{r,it}$$

is integrated of order $\theta = \max\{\delta_{max}, \vartheta_{max}, d_{max}\}$, or $\bar{\bar{y}} \sim I(\theta)$, so long as $\lambda_{r,i} \neq 0$ and $\mu_{r,i} \neq 0$ for any i, r. For example, when $\delta_{max} > \vartheta_{max} > d_{max}$, we have $\bar{\bar{y}} \sim I(\delta_{max})$ and the cross-sectionally and regionally averaged error term, i.e. $\bar{\bar{\varepsilon}} = O_p(N^{-1/2}(1 + T^{d_{max}-1/2}))$, is dominated by G_t and $F_{r,t}$ as $N \to \infty$ and in that case the estimate of θ, say $\hat{\theta}$, estimates δ_{max}. In our estimation method, we do not impose such a restriction to keep things general. Using the exact local Whittle estimation by Shimotsu (2010), the asymptotic behavior of $\hat{\theta}$ is shown in the following result.

Theorem 2.1. *Under Assumptions A-E, as $T \to \infty$,*

$$m^{1/2}(\hat{\theta} - \theta) \to_d N(0, 1/4).$$

The convergence rate depends on the bandwidth parameter, m, which satisfies Assumption A.4. This result parallels the ones by Shimotsu (2010) who considers a single time series employing similar conditions.

To estimate the global and regional factors, we prewhiten the series by $\hat{\theta}$ and write the prewhitened version of (1) as

$$y_{r,it}(\hat{\theta}) = \mu'_{r,i} G_t(\hat{\theta}) + \lambda'_{r,i} F_{r,t}(\hat{\theta}) + \epsilon_{r,it}(\hat{\theta})$$

which can be written in matrix notation based on (2) as

$$y_t(\hat{\theta}) = \Lambda^* \ F_t^*(\hat{\theta}) + \epsilon_t(\hat{\theta}).$$

We then use the sequential least squares (SLS) procedure that is proposed by Breitung and Eickmeier (2016). We outline the steps of such an algorithm in which the main goal is to minimize the residual sum of squares (RSS) function

$$\begin{aligned} S\left(F^*(\hat{\theta}), \Lambda^*(\hat{\theta})\right) &= \sum_{t=1}^{T} \left(y_t(\hat{\theta}) - \Lambda^* F_t^*(\hat{\theta})\right)' \left(y_t(\hat{\theta}) - \Lambda^* F_t^*(\hat{\theta})\right) \\ &= \sum_{r=1}^{R} \sum_{i=1}^{N_r} \sum_{t=1}^{T} \left(y_{r,it}(\hat{\theta}) - \gamma'_{r,i} G_t(\hat{\theta}) - \lambda'_{r,i} F_{r,t}(\hat{\theta})\right)^2 \end{aligned} \tag{4}$$

by a sequence of two least-squares regressions until RSS achieves a minimum. The algorithm is easily executed as follows:

1. The algorithm is initialized by using initial estimators of the global and regional factors, $\hat{G}^{(0)}(\hat{\theta}) = \left(\hat{G}_1^{(0)}(\hat{\theta}), \ldots, \hat{G}_T^{(0)}(\hat{\theta})\right)'$ and $\hat{F}_r^{(0)}(\hat{\theta}) = \left(\hat{F}_{r,1}^{(0)}(\hat{\theta}), \ldots, \hat{F}_{r,T}^{(0)}(\hat{\theta})\right)'$. Such estimators can be obtained by canonical correlation analysis (CCA).

2. Once initial estimators are obtained, the corresponding factor loadings at the initial step are estimated from the time-series regression $y_{r,it} = \mu'_{r,i} \hat{G}_t^{(0)}(\hat{\theta}) +$

$\lambda'_{r,i} F^{(0)}_{r,t}(\hat{\theta}) + \widetilde{\epsilon}_{r,it}(\hat{\theta})$ that construct the factor loadings matrix, $\hat{\Lambda}^{*(0)}$, as in Equation (2).

3. The global and regional factors in the next step, $\hat{G}^{(1)}(\hat{\theta})$ and $\hat{F}^{(1)}_{r,1}(\hat{\theta})$, are updated from the least-squares regression of $y_t(\hat{\theta})$ on $\hat{\Lambda}^{*(0)}$ to obtain $F^{*(1)}_t(\hat{\theta}) = \left(\hat{\Lambda}^{*(0)\prime}\hat{\Lambda}^{*(0)}\right)^{-1} \hat{\Lambda}^{*(0)\prime} y_t(\hat{\theta})$.

4. Next, the updated factors $F^{*(1)}_t(\hat{\theta})$ are used to get the associated factor loading matrix, $\hat{\Lambda}^{*(1)}$, as in step 2.

5. Steps 3 and 4 are repeated until RSS converges to a minimum from which $\hat{F}^{*}(\hat{\theta})$ and $\hat{\Lambda}^{*}$ are collected.

CCA is a standard tool in multivariate statistics and is a way of measuring the linear relationship between two multidimensional variables. Such an analysis finds two sets of basis vectors, one for each variable, such that the correlations between the projections of the variables onto these bases are mutually maximized. Along this line, Breitung and Pigorsch (2013) propose a canonical correlation approach to estimate the number of dynamic factors in a dynamic factor model.

In a multilevel setup, Breitung and Eickmeier (2016) and Choi et al. (2016) propose CCA to get the initial estimates of the global and regional factors to ensure that the procedure listed above starts in a suitable vicinity of the global minimum. CCA is carried out in 2 steps. At the first step, in each region, $\mathbf{r} = \mathbf{r}_G + \mathbf{r}_{F_r}$ principal components are estimated obtaining R consistent factor spaces of the form $\hat{F}^{+}_{r,t}(\hat{\theta})$ for $r = 1, \ldots, R$ which will eventually share a common component yielding the initial global factor after the second step. Let $\mathscr{H}_t(\hat{\theta}) = \left(\hat{F}^{+}_{r,t}(\hat{\theta}), \hat{F}^{+}_{s,t}(\hat{\theta})\right)'$ with $c^0\mathscr{H}_t(\hat{\theta})$ denoting the canonical variables, the CCA (at the second step of the procedure) solves the following maximization problem:

$$\max \left\{ c^0 \Sigma_{01} c^1 / \left[c^0 \Sigma_{00} c^0 \cdot c^1 \Sigma_{11} c^1 \right]^{1/2} \right\}$$

$$\text{s.t. } c^0 \Sigma_{00} c^0 = 1, \text{ and } c^1 \Sigma_{11} c^1 = 1,$$

where $\Sigma_{00} = \text{var}\left(\mathscr{H}_t(\hat{\theta})\right)$, $\Sigma_{11} = \text{var}\left(\mathscr{H}_{t-1}(\hat{\theta})\right)$ and, $\Sigma_{01} = \text{cov}\left(\mathscr{H}_t(\hat{\theta}), \mathscr{H}_{t-1}(\hat{\theta})\right)$. The resulting linear combination with the largest canonical correlation will be the estimate of the global factor, $\hat{G}^{(0)}(\hat{\theta})$. Subsequently, we regress original principal components of the region r, $\hat{F}^{+}_{r,t}(\hat{\theta})$, on the estimated global factors in order to find $\hat{F}^{(0)}_r(\hat{\theta})$ for all $r = 1, \ldots, R$.

As pointed out earlier, Λ^{*0} and $F_t^{*0}(\hat{\theta})$ are not separately identifiable. In order to identify the common component $\xi_t^*(\hat{\theta}) = \Lambda^* F_t^*(\hat{\theta})$ we choose the nonsingular matrix Q in (3) to preserve identification of the factors. We use the standard normalizations in PC analysis given in Assumption E. Note that even when global and regional factors are orthogonal to each other, the correlation between regional factors from different regions are allowed as discussed before. We follow the steps proposed by Breitung and Eickmeier (2016) to adapt to the normalization. Such steps consist first of regressing the regional factors $\hat{F}_{r,t}(\hat{\theta})$ on $\hat{G}_t(\hat{\theta})$ in order to get the orthogonalized regional factors, and second, extracting the normalized global and regional factors after running PC analysis of the respective common component.

The obtained factor estimates span the factor space up to a rotation where the transpose of the rotation matrix, H_{F^*}, is defined as

$$H_F' = \hat{V}' \left(\hat{F}^*(\theta)' F^*(\theta)/T \right) (\Lambda^{*'} \Lambda^*/N)$$

with $\hat{V} = \hat{\Lambda}^{*'} \hat{\Lambda}^*/N$ where $\hat{\Lambda}^*$ is a consistent estimate of Λ^* whose convergence rate is at least $m^{1/2}$.

Define for a fixed t,

$$\Gamma_t^*(\theta) = E \left(\Lambda^{*'} \epsilon_t(\theta) \epsilon_t(\theta)' \Lambda^* \right)$$

and

$$\Sigma_{\Lambda^*} = E \left(\Lambda^{*'} \Lambda^* \right).$$

Then we establish the asymptotic behavior of the prewhitened factor estimates in the following result.

Theorem 2.2. *Under Assumptions A-E if $N/m \to 0$ as $(N,T)_j \to \infty$,*

$$\sqrt{N}(\hat{F}_t^*(\hat{\theta}) - F_t^*(\theta) H_{F^*}) \to_d N(0, \Sigma_{\Lambda^*}^{-1} \Gamma_t(\theta) \Sigma_{\Lambda^*}^{-1})$$

for a fixed t.

The variance-covariance matrix components can be separately estimated along the lines of Bai and Ng (2006). The rate requirement $Nm^{-1} \to 0$ is imposed to remove the errors due to the estimation of the prewhitening parameter θ

at the first step and requires a larger time series in the estimation than would be needed if a parametric approach were to be used.

After obtaining the prewhitened factor estimates, the original factor estimates can recovered by integrating them back by $\hat{\theta}$ as

$$\Delta^{-\hat{\theta}} \hat{F}_t^*(\hat{\theta}) = \hat{F}_t^*$$

omitting dependence on $\hat{\theta}$ and assuming away the initial conditions that are negligible under joint asymptotics, see Ergemen and Velasco (2017).

Using the original factor estimates, integration orders of the global and regional factors can be estimated either parametrically based on a conditional-sum-of-squares (CSS) criterion, see e.g. Ergemen and Velasco (2017), or semiparametrically, see e.g. Abadir et al. (2007).

The factor loadings matrix Λ^* can also be consistently estimated by a least-squares method since the factors $\hat{F}_t^*$ are observable. This enables the recovery of the estimated regression residuals,

$$\hat{\epsilon}_t = y_t - \hat{\Lambda}^* \hat{F}_t^*$$

from which the residual memory parameters $d_{r,i0}$ can be estimated again either parametrically based on a CSS criterion, see e.g. Ergemen and Velasco (2017) or semiparametrically based on a local Whittle method, such as that of Abadir et al. (2007).

3 Determining the number of regional and global factors

The model in (1) assumes that the number of factors in each region and the number of global factors is fixed and known. Although a crucial step in the identification of the model is to accurately estimate the numbers of such factors, most of the empirical literature fixes the number of regional and global factors to be one or analyze directly some alternative models considering more factors without using formal information criteria.

Although there are many methodologies to estimate the number of the static factors in one-level factor models, see e.g. Bai and Ng (2002), Alessi et al. (2010), Onatski (2010), Kapetanios (2010) and Ahn and Horenstein (2013), a formal methodology to estimate the number of static factors in a multi-level factor model is not yet available to the best of our knowledge. The only exception is the proposal of Hallin and Liška (2011) who allow for identifying and estimating joint and block-specific common factors in the context of dynamic factor models.

Inspired by the methodology of Hallin and Liška (2011) and using the well known information criteria of Bai and Ng (2002), we propose a new procedure for identifying the number of regional and global factors under our setup. We retain Assumptions A-E imposed to study the model in (1). Our assumptions are in line with those of Bai and Ng (2002) but we have extra conditions pertaining to the long-range dependence of factors and the idiosyncratic terms.

Bai and Ng (2002) consider an approximate static factor model and suggest a penalty criteria function of the form

$$PC(k) = V\left(k, \hat{F}^k\right) + \ k\, g(N, T), \tag{5}$$

where $V\left(k, \hat{F}^k\right)$ is the sum of squared residuals when k factors are estimated. The essence of the criteria is to find penalty functions, $g(N, T)$, which can consistently estimate the number of static factors. Assuming that there exists a bounded and positive integer k_{max} number of static factors such that $r \leq k_{max}$, Theorem 2 and Corollary 1 in Bai and Ng (2002) provide necessary conditions to consistently estimate the number of static factors r. Bai and Ng (2002) provide six choices for the penalty function and indicate the corresponding criteria as PC1, PC2, PC3, IC1, IC2, and IC3. Criteria IC1 and IC2 are more often used in empirical applications in the literature.

Because there is no available literature regarding the estimation of the number of factors in presence of long memory, we first focus on a type-II fractionally integrated single-level factor model to discuss how to estimate the number of factors, s, in a static factor model with long memory by using the same information criteria provided by Bai and Ng (2002). Consider the model (1) with $R = 1$, that is

$$y_{it} = \lambda_i' F_t + \epsilon_{it},$$

where $F_t = \Delta_t^{-\vartheta_0} v_t$, and $\epsilon_{it} = \Delta_t^{-d_{i0}} u_{r,it}$ as before. Let $\varsigma_{i0} \equiv \max\left(\vartheta_0, d_{i0}\right)$, then fractional differencing each y_i by ς_{i0} we can consistently apply information criteria of Bai and Ng (2002) regardless of the values of ϑ_0 or d_{i0}, otherwise the estimation of r will be dramatically affected when $d_{i0} > 0$. The latter is implied by Assumption C in Bai and Ng (2002). Once fractional differencing is applied, the consistency proof in Theorem 2 by Bai and Ng (2002) is still valid.[1] When $\varsigma_0 \leq 1$, taking first differences would be enough to ensure the consistency of number of factors estimate. Nevertheless, we suggest estimating first ς by using the Extended Local Whittle method (ELW) of Abadir et al. (2007), which covers the stationary and nonstationary regions even beyond the unit root, in order to get $\hat{\varsigma}$ and later, fractional differencing each y_i by $\hat{\varsigma}$.

Table 1 reports a small Monte Carlo simulation to illustrate this methodology. As can be seen, Case 1 exemplifies that when the residual fractional memory is stationary ($d_i = 0.4$), the information criteria IC1-IC3 and PC1-PC3 perform well regardless of whether we neglect the memory or if we take the first difference in y_{it} or if we fractional differencing by $\varsigma = 0.4$ in contrast to Case 2 where $d_i = 1$. In such a case, information criteria are not useful anymore when neglecting memory but continue to work well when taking first differences in observables. Cases 3 and 4 are two examples where we do not have a cointegrating relationship between y_{it} and F_t. Even in absence of cointegration, information criteria perform well when fractional differencing by ς. Naturally it is enough to take only the first difference in Case 3 since $\varsigma < 1$.

We now extend the aforementioned methodology to allow for more than one-level in a factor model. Our methodology to estimate the number of static regional and global factors adopts the method of Hallin and Liška (2011) which identifies and estimates joint and block-specific common factors by using the identification method of Hallin and Liška (2007) in the nature of dynamic factor models. In the case of Hallin and Liška (2011), their joint common factors may be interpreted as a global or pervasive top-level factor in our case whereas the block-specific factor would be the regional or non-pervasive sub-level factor.

For the sake of simplicity, consider only two regions or blocks (B_x, B_y) in model 1, only one regional factor in each region and one global factor. We can now divide our data into three different factor spaces. Call the marginal factor spaces as those two different spaces spanned by the individual blocks of data B_x and B_y and call the joint factor space as that spanned by the complete block $B_{x\cup y}$. In these spaces, we see that $s_{B_x} = 2$, $s_{B_y} = 2$, and $s_{Bx\cup y} = 3$ given that we have only one regional factor in each region and only one global factor. The latter means that both marginal factor spaces consist of two static factors whereas the joint factor space consists of three static factors. The number of factors in each one of these three

[1] As a matter of fact, it would be only necessary to fractionally difference by d_{i0} to consistently estimate r. However d_{i0} remains unknown until after the model is estimated.

Table 1: Number of static factors under negligence of long memory, first differencing, and fractional differencing ($N = 40$, $T = 300$, $s = 3$ and $k_{max} = 10$).

	Neglected memory	First Diff.	Diff. by ς	Neglected memory	First Diff.	Diff. by ς
	Case 1. $d_i = 0.4\ \forall i$		$\vartheta = 0.8$	Case 2. $d_i = 1\ \forall i$		$\vartheta = 1.5$
IC1	3	3	3	10	3	3
IC2	3	3	3	10	3	3
IC3	3	3	3	10	3	3
PC1	5	3	3	10	3	3
PC2	5	3	3	10	3	3
PC3	6	3	3	10	3	3
	Case 3. $d_i = 0.8\ \forall i$		$\vartheta = 0.4$	Case 4. $d_i = 1.5\ \forall i$		$\vartheta = 1$
IC1	10	3	3	10	6	3
IC2	10	3	3	10	5	3
IC3	10	3	3	10	8	3
PC1	10	3	3	10	8	3
PC2	10	3	3	10	8	3
PC3	10	3	3	10	8	3

Notes: The DGP is $y_{it} = \lambda_i' F_t + \Delta_t^{-d_{i0}} \epsilon_{it}$. $\epsilon_{it} \overset{iid}{\sim} N(0,1)$ are generated independently. $F_t = \Delta_t^{-\vartheta} v_t$ with $v_t \sim IIDN(0,1)$. The IC1, IC2, IC3, PC1, PC2, and PC3 information criteria of Bai and Ng (2002) are used to estimate the number of factors. The average numbers of factors across the replications are presented. We compare three cases: i) when memory is neglected, ii) taking the first difference on y_{it}, and iii) fractional differencing y_{it} with $\varsigma_0 = max(\vartheta_0, d_0)$. All experiments are based on 1000 replications.

factor spaces is consistently estimated by using the information criteria in Bai and Ng (2002) after fractional differencing by $\hat{\varsigma}$ as discussed before. Theorem 2 in Bai and Ng (2002) is still valid after fractional differencing. It is worth mentioning that Alessi et al. (2010) introduce in the penalty function a new parameter in order to avoid that the number of factors can be overestimated or underestimated as with the information criteria of Bai and Ng (2002) when the variables are heteroskedastic, for instance, in some empirical applications as in this paper.

The simple Venn diagram in Figure 1 displays the strategy discussed above for the case of 2 blocks. Green sector represents the part of the factor space which is shared by both regions and consists of one static factor (the global factor). The marginal factor space B_x is represented by blue + green sectors having two static factors whereas the marginal factor space B_y is the yellow + green sectors and also has two factors. Naturally, the number of regional static factors is directly obtained after computing the number of global factor by the inclusion-exclusion principle, i.e. $s_{B_x \cup B_y} = s_{B_x} + s_{B_y} - s_{B_x \cap B_y}$, from which we would get $s_{B_x \cap B_y} = 1$ (the global factor).

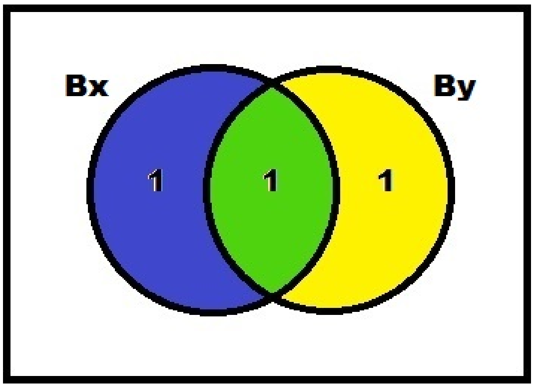

Figure 1: Representation of the three factor spaces spanned by two regions or blocks of data.

The complexity of this methodology increases in the number of regions. Clearly, when we have R regions, the number of blocks to be analyzed will be the power set minus one, $2^R - 1$. Furthermore, we should compute the number of global factors by using each one of the number of factors (cardinalities) estimated in the individual, pairwise, triple-wise, etc. sets by the inclusion-exclusion principle. The number of regional factors in region R would be determined by subtracting the number of factors previously estimated in each one of the intersections where region R interacts with the number of factors previously estimated only in the region R.

As an example, consider now three regions, we have the following blocks: B_x, B_y, B_z, $B_{x\cup y}$, $B_{x\cup z}$, $B_{y\cup z}$, and $B_{x\cup y\cup z}$ from which after fractional differencing all the variables in the data set, we compute the number of factors which span each one of the seven blocks. The number of factors in pair-wise blocks will be given by $s_{B_{x\cap y}} = s_{B_x} + s_{B_y} - s_{B_{x\cup y}}$, for instance. The global factor will be given by $s_{B_{x\cap y\cap z}} = s_{B_{x\cup y\cup z}} - s_{B_x} - s_{B_y} - s_{B_z} + s_{B_{x\cap y}} + s_{B_{x\cap z}} + s_{B_{y\cap z}}$ and the number of regional factors of the region x by $s_{B_x} - s_{B_{x\cap y\cap z}} - s_{B_{x\cap y}} - s_{B_{x\cap z}}$, for instance.

Naturally, with this methodology it is possible to specify not only the number of factors corresponding to the global and regional levels but also the number of factors in each one of the three pairwise blocks of regions.

4 Finite sample properties

In this section we study the finite-sample properties of the sequential least squares (SLS) procedure to investigate the performance of the model in (1) and the methodology proposed to estimate the number of global and regional factors.[2]

4.1 Two-level factor model

We first present four Monte Carlo studies to study the performance of our model. In our simulation studies we are generating a fractional cointegration relationship between $y_{r,it}$ and the global factor (G_t) since we believe such a relationship is likely in several empirical studies.

In the first Monte Carlo study, whose results are presented in Tables 2, 3, and 4, we analyze the performance of our model with $R = 2$, $N_r \in \{20, 80\}$, and different sample sizes with $T = \{150, 1000, 5000\}$, respectively. One global factor and one regional factor in each region are considered for simplicity although more factors are allowed. The global, both regional factors, and all the idiosyncratic terms are independently generated by ARFIMA(1,d*,0) processes where d* corresponds to δ, ϑ_r or $d_{i,r}$ as appropriate. Autoregressive parameters are 0.5 for the unobservable factors and 0.1 for the idiosyncratic errors, following Breitung and Eickmeier (2016). Working $u_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ with ϕ controlling the signal-to-noise-ratio with $\phi = \{5, 2, 0.5\}$, corresponding to low, medium, and high signal-

[2] We are deeply grateful to Sandra Eickmeier for sharing her Matlab code with us.

to-noise-ratios. $w_t \overset{iid}{\sim} N(0, \sigma_w)$ and $v_{r,t} \overset{iid}{\sim} N(0, \sigma_{v_r})$ controlling the ratio $\frac{\sigma_{v_r}}{\sigma_w}$ to study the relative impact of the factors to each other. Furthermore, all factor loadings are generated as $N(1,1)$, following Boivin and Ng (2006). All results are based on 1000 replications of the model.

For each experiment, after collecting the estimated regional and global factors, we estimate the memory parameters $\hat{\vartheta}_r$, $\hat{\delta}$ using the Extended Local Whittle procedure. We also regress the actual factors (global or regional) on the estimated ones in order to study the reliability of the procedure by computing coefficient of determinations of the global and regional factors, denoted as R_G^2 and $R_{R_r}^2$, respectively. Both coefficients can be considered as a measure of consistency for all t, see Bai (2003). Finally, $\bar{\hat{d}}_{R_r}$ denotes the average of the estimated residual integration orders in the region r by the CSS procedure of Ergemen and Velasco (2017).

Memory estimates of the global and regional factors as well as those of the residuals are accurately estimated no matter the sample size or the persistence levels in $d_{r,i0}$. Changes in the level of the signal-to-noise-ratios do not affect the estimated residual integration orders. On the other hand, even when $d_{r,i0} = 0$, the global and regional factors are consistently estimated and as long as $d_{r,i0} < 0.5$, the accuracy of the global and regional factors is not significantly distorted. This suggests that practitioners can estimate the model in (1) without taking fractional differences or first differences in the variables and if $\hat{d}_{r,i} < 0.5$ their global and regional factors will indeed be accurately estimated. Cases for $\hat{d}_{r,i} \geq 0.5$, are discussed in the fourth Monte Carlo simulation. Finally, a low signal-to-noise-ratio makes the regional factors less precisely estimated. Note that such ratios do not affect the accuracy of the estimated global factor. Our findings indicate that it is possible that the use of the canonical correlation procedures is sufficiently robust for specifying the global factor. Using CCA to estimate the number of dynamic factors, Breitung and Pigorsch (2013) point out that CCA is useful for a wide range of stationary or mixing processes and particularly works better than usual PC methods if the variances of the factors are very different.

In the second Monte Carlo study, for which the results are presented in Table 5, we find that the performance of the model is not affected by increasing the number of regions or varying the persistence of the regional factors. Factors, the idiosyncratic terms, and loading factors are generated as before. We now study four regions ($R = 4$) with different persistence levels. We only consider 20 variables ($N_r = 20$) in each region r. All the standard deviations are simpler (0.5,1, and 2). Conclusions are similar to those in the first simulation study.

Table 2: First Monte Carlo simulation with $T = 150$, $N_r \in (20, 80)$, and $R = 2$.

			$N_r = 20$									$N_r = 80$							
ϕ	σ_{v_r}	σ_w	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$		$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$
									$\vartheta_{r0} = 0.4$	$\delta_0 = 0.6$		$d_{r,i0} = 0$							
0.5	2	1	0.429	0.426	0.642	0.994	0.967	0.965	0.044	0.045		0.434	0.439	0.642	0.999	0.966	0.969	0.044	0.044
0.5	1	1	0.428	0.432	0.633	0.994	0.951	0.954	0.044	0.045		0.432	0.432	0.646	0.999	0.966	0.968	0.044	0.044
0.5	1	2	0.422	0.427	0.644	0.999	0.951	0.953	0.045	0.044		0.431	0.437	0.644	1.000	0.968	0.966	0.044	0.045
2	2	1	0.427	0.438	0.628	0.975	0.952	0.955	0.044	0.044		0.426	0.435	0.641	0.994	0.966	0.966	0.045	0.044
2	1	1	0.410	0.414	0.624	0.975	0.904	0.904	0.045	0.044		0.430	0.431	0.634	0.994	0.955	0.954	0.044	0.045
2	1	2	0.414	0.412	0.635	0.994	0.905	0.905	0.045	0.044		0.427	0.431	0.641	0.999	0.954	0.952	0.044	0.044
5	2	1	0.387	0.384	0.633	0.984	0.815	0.816	0.046	0.045		0.422	0.423	0.641	0.996	0.931	0.931	0.045	0.044
5	1	1	0.379	0.385	0.633	0.984	0.818	0.817	0.045	0.044		0.429	0.421	0.639	0.996	0.930	0.929	0.045	0.044
5	1	2	0.388	0.383	0.636	0.984	0.818	0.818	0.046	0.045		0.418	0.423	0.646	0.996	0.930	0.930	0.045	0.045
									$\vartheta_{r0} = 0.6$	$\delta_0 = 1$		$d_{r,i0} = 0.25$							
0.5	2	1	0.637	0.642	1.023	0.999	0.914	0.913	0.235	0.234		0.640	0.637	1.027	1.000	0.915	0.921	0.236	0.236
0.5	1	1	0.636	0.634	1.026	0.999	0.908	0.906	0.235	0.235		0.643	0.640	1.028	1.000	0.912	0.913	0.236	0.235
0.5	1	2	0.636	0.639	1.029	0.999	0.907	0.906	0.235	0.235		0.640	0.640	1.024	1.000	0.911	0.915	0.236	0.236
2	2	1	0.632	0.638	1.012	0.995	0.909	0.908	0.235	0.236		0.632	0.638	1.028	0.999	0.915	0.915	0.236	0.236
2	1	1	0.618	0.626	1.014	0.995	0.869	0.873	0.235	0.235		0.634	0.639	1.025	0.999	0.907	0.904	0.236	0.236
2	1	2	0.618	0.623	1.021	0.999	0.872	0.870	0.236	0.235		0.635	0.640	1.032	1.000	0.902	0.906	0.235	0.235
5	2	1	0.597	0.593	1.021	0.997	0.816	0.818	0.235	0.235		0.635	0.633	1.023	0.999	0.888	0.886	0.235	0.236
5	1	1	0.599	0.597	1.018	0.997	0.815	0.813	0.235	0.235		0.633	0.631	1.026	0.999	0.889	0.890	0.235	0.236
5	1	2	0.593	0.599	1.018	0.997	0.820	0.814	0.235	0.235		0.637	0.641	1.028	0.999	0.891	0.891	0.235	0.235
									$\vartheta_{r0} = 0.8$	$\delta_0 = 1$		$d_{r,i0} = 0.45$							
0.5	2	1	0.839	0.838	1.031	0.998	0.826	0.827	0.441	0.440		0.835	0.837	1.031	1.000	0.822	0.820	0.440	0.441
0.5	1	1	0.835	0.838	1.029	0.998	0.821	0.808	0.440	0.440		0.842	0.844	1.032	1.000	0.815	0.821	0.440	0.441
0.5	1	2	0.835	0.833	1.028	0.998	0.820	0.821	0.441	0.441		0.840	0.842	1.029	1.000	0.825	0.825	0.440	0.440
2	2	1	0.833	0.838	1.017	0.993	0.822	0.813	0.440	0.440		0.848	0.841	1.033	0.998	0.822	0.833	0.441	0.440
2	1	1	0.829	0.822	1.018	0.993	0.799	0.808	0.439	0.441		0.836	0.838	1.031	0.998	0.820	0.812	0.441	0.441
2	1	2	0.822	0.830	1.029	0.998	0.804	0.809	0.440	0.440		0.838	0.837	1.022	1.000	0.819	0.814	0.440	0.441
5	2	1	0.804	0.804	1.016	0.995	0.748	0.743	0.440	0.440		0.845	0.849	1.027	0.999	0.806	0.801	0.441	0.441
5	1	1	0.810	0.806	1.023	0.996	0.760	0.758	0.441	0.441		0.847	0.843	1.025	0.999	0.804	0.813	0.441	0.441
5	1	2	0.804	0.805	1.018	0.996	0.744	0.760	0.441	0.439		0.845	0.846	1.037	0.999	0.787	0.813	0.440	0.441

Notes: The averages of the memory estimated of the unobservable factors in both levels, the residual memory estimates, and the measure of consistency of the unobservable factors estimated are presented in the report. The DGP is $y_{r,it} = \mu'_{r,i} G_t + \lambda'_{r,i} F_{r,t} + \Delta_t^{-d_{r,i0}} \epsilon_{r,it}$. $r = 1, 2$ and $i \in (20, 80)$ and $t = 150$. $\epsilon_{r,it} \overset{iid}{\sim} N(0, 2\phi)$ are generated independently with ϕ controlling the signal-to-noise-ratio with $\phi = \{5, 2, 0.5\}$. Only one top-level factor and only one block-specific factor in each block are considered. $G_t = 0.5G_{t-1} + \Delta_t^{-\delta} w_t$ with $w_t^G \sim IIDN(0, \sigma_w)$ and $\sigma_w \in (1, 2)$. $F_{r,t} = 0.5F_{r,t-1} + \Delta_t^{-\vartheta_r} v_t$ with $v_t \sim IIDN(0, \sigma_v)$ and $\sigma_v \in (1, 2)$. $\hat{\vartheta}_r$ and $\hat{\delta}$ are the memory estimated of the factors by Extended Local Whittle procedure. R^2_G and $R^2_{R_r}$ are the R^2 of a regression of actual on estimates global and regionals factors, respectively. $\bar{\hat{d}}_{R1}$ and $\bar{\hat{d}}_{R2}$ are the average of the residual memory estimates. All experiments are based on 1000 replications.

Table 3: First Monte Carlo simulation with $T = 1000$, $N_r \in (20, 80)$, and $R = 2$.

			$N_r = 20$								$N_r = 80$							
ϕ	σ_{v_r}	σ_w	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R1}$
								$\vartheta_{r0} = 0.3$		$\delta_0 = 0.5$		$d_{r,i0} = 0$						
0.5	2	1	0.304	0.306	0.505	0.995	0.989	0.988	0.062	0.062	0.304	0.302	0.506	0.999	0.992	0.992	0.062	0.062
0.5	1	1	0.298	0.300	0.507	0.995	0.975	0.976	0.062	0.062	0.305	0.303	0.508	0.999	0.989	0.989	0.062	0.062
0.5	1	2	0.301	0.300	0.508	0.999	0.976	0.977	0.062	0.061	0.303	0.303	0.508	1.000	0.988	0.989	0.062	0.062
2	2	1	0.300	0.303	0.500	0.982	0.977	0.977	0.062	0.062	0.302	0.301	0.505	0.996	0.989	0.989	0.062	0.062
2	1	1	0.294	0.295	0.506	0.982	0.930	0.930	0.062	0.061	0.301	0.301	0.509	0.996	0.978	0.977	0.062	0.062
2	1	2	0.294	0.294	0.508	0.995	0.930	0.930	0.061	0.062	0.301	0.304	0.505	0.999	0.977	0.978	0.062	0.062
5	2	1	0.283	0.284	0.503	0.989	0.851	0.851	0.062	0.062	0.299	0.298	0.505	0.997	0.956	0.956	0.062	0.062
5	1	1	0.287	0.281	0.507	0.989	0.854	0.852	0.062	0.061	0.296	0.300	0.504	0.997	0.956	0.956	0.062	0.062
5	1	2	0.284	0.283	0.503	0.989	0.853	0.850	0.062	0.061	0.299	0.297	0.507	0.997	0.956	0.956	0.062	0.062
								$\vartheta_r 0 = 0.5$		$\delta_0 = 0.9$		$d_{r,i0} = 0.25$						
0.5	2	1	0.508	0.506	0.901	1.000	0.953	0.949	0.261	0.261	0.508	0.510	0.904	1.000	0.954	0.953	0.261	0.261
0.5	1	1	0.507	0.508	0.903	1.000	0.945	0.945	0.261	0.261	0.510	0.509	0.903	1.000	0.949	0.949	0.261	0.261
0.5	1	2	0.509	0.508	0.903	1.000	0.945	0.943	0.261	0.261	0.507	0.508	0.901	1.000	0.953	0.954	0.261	0.261
2	2	1	0.506	0.506	0.900	0.999	0.944	0.950	0.260	0.261	0.510	0.509	0.902	1.000	0.952	0.952	0.261	0.261
2	1	1	0.502	0.499	0.898	0.999	0.918	0.923	0.261	0.261	0.508	0.510	0.903	1.000	0.948	0.946	0.260	0.261
2	1	2	0.503	0.502	0.900	1.000	0.922	0.923	0.261	0.261	0.506	0.505	0.903	1.000	0.946	0.945	0.261	0.261
5	2	1	0.491	0.490	0.899	0.999	0.874	0.874	0.261	0.261	0.501	0.506	0.901	1.000	0.935	0.935	0.261	0.261
5	1	1	0.491	0.491	0.900	0.999	0.875	0.874	0.261	0.260	0.503	0.505	0.901	1.000	0.936	0.930	0.261	0.261
5	1	2	0.491	0.492	0.900	0.999	0.873	0.874	0.261	0.261	0.504	0.507	0.904	1.000	0.938	0.934	0.261	0.261
								$\vartheta_r 0 = 0.7$		$\delta_0 = 0.9$		$d_{r,i0} = 0.45$						
0.5	2	1	0.713	0.713	0.902	0.999	0.863	0.868	0.465	0.464	0.715	0.717	0.905	1.000	0.869	0.861	0.464	0.465
0.5	1	1	0.710	0.710	0.901	1.000	0.862	0.855	0.464	0.464	0.715	0.715	0.902	1.000	0.862	0.868	0.464	0.464
0.5	1	2	0.712	0.712	0.904	1.000	0.862	0.862	0.465	0.465	0.714	0.714	0.903	1.000	0.868	0.858	0.465	0.464
2	2	1	0.713	0.710	0.897	0.998	0.862	0.852	0.464	0.464	0.715	0.717	0.903	1.000	0.860	0.866	0.465	0.464
2	1	1	0.707	0.706	0.897	0.998	0.846	0.846	0.465	0.465	0.714	0.713	0.902	1.000	0.856	0.866	0.465	0.465
2	1	2	0.708	0.708	0.902	0.999	0.849	0.850	0.465	0.464	0.714	0.712	0.903	1.000	0.858	0.860	0.464	0.465
5	2	1	0.697	0.700	0.896	0.999	0.825	0.818	0.465	0.464	0.713	0.713	0.904	1.000	0.857	0.852	0.465	0.465
5	1	1	0.698	0.698	0.900	0.999	0.827	0.827	0.465	0.465	0.709	0.711	0.903	1.000	0.849	0.849	0.465	0.465
5	1	2	0.697	0.696	0.898	0.999	0.824	0.821	0.464	0.464	0.713	0.712	0.900	1.000	0.862	0.848	0.465	0.464

Notes: The averages of the memory estimated of the unobservable factors in both levels, the residual memory estimates, and the measure of consistency of the unobservable factors estimated are presented in the report. The DGP is the same as that of Table 2, except that now $T = 1000$. All experiments are based on 1000 replications.

Table 4: First Monte Carlo simulation with $T = 5000$, $N_r \in (20, 80)$, and $R = 2$.

						$N_r = 20$									$N_r = 80$			
ϕ	σ_{v_r}	σ_w	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R1}$
								$\vartheta_{r0} = 0.24$		$\delta_0 = 0.44$		$d_{r,i0} = 0$						
0.5	2	1	0.246	0.246	0.447	0.996	0.994	0.994	0.064	0.064	0.246	0.243	0.449	0.999	0.997	0.997	0.064	0.064
0.5	1	1	0.246	0.244	0.446	0.996	0.982	0.982	0.064	0.064	0.246	0.246	0.448	0.999	0.994	0.994	0.064	0.064
0.5	1	2	0.244	0.245	0.448	0.999	0.982	0.982	0.064	0.064	0.246	0.247	0.449	1.000	0.994	0.994	0.064	0.064
2	2	1	0.244	0.244	0.446	0.984	0.982	0.982	0.064	0.064	0.245	0.246	0.448	0.996	0.994	0.994	0.064	0.064
2	1	1	0.241	0.242	0.446	0.984	0.938	0.939	0.064	0.064	0.246	0.244	0.447	0.996	0.983	0.983	0.064	0.064
2	1	2	0.241	0.240	0.447	0.996	0.937	0.937	0.064	0.064	0.244	0.244	0.449	0.999	0.983	0.983	0.064	0.064
5	2	1	0.234	0.234	0.447	0.990	0.861	0.859	0.064	0.064	0.243	0.244	0.448	0.998	0.962	0.962	0.064	0.064
5	1	1	0.236	0.236	0.448	0.990	0.859	0.859	0.064	0.064	0.244	0.243	0.448	0.998	0.962	0.962	0.064	0.064
5	1	2	0.235	0.235	0.446	0.990	0.860	0.860	0.064	0.064	0.243	0.242	0.446	0.998	0.962	0.962	0.064	0.064
								$\vartheta_r 0 = 0.44$		$\delta_0 = 0.84$		$d_{r,i0} = 0.25$						
0.5	2	1	0.449	0.450	0.845	1.000	0.967	0.969	0.264	0.264	0.449	0.448	0.846	1.000	0.971	0.970	0.264	0.264
0.5	1	1	0.448	0.449	0.843	1.000	0.963	0.963	0.264	0.264	0.450	0.449	0.847	1.000	0.967	0.971	0.264	0.264
0.5	1	2	0.448	0.447	0.845	1.000	0.960	0.960	0.264	0.264	0.448	0.449	0.845	1.000	0.969	0.968	0.264	0.264
2	2	1	0.448	0.448	0.846	1.000	0.961	0.963	0.264	0.264	0.448	0.448	0.845	1.000	0.970	0.970	0.264	0.264
2	1	1	0.444	0.443	0.843	1.000	0.940	0.939	0.264	0.264	0.448	0.449	0.845	1.000	0.964	0.964	0.264	0.264
2	1	2	0.444	0.444	0.845	1.000	0.941	0.941	0.264	0.264	0.448	0.448	0.845	1.000	0.963	0.965	0.264	0.264
5	2	1	0.438	0.438	0.844	1.000	0.899	0.895	0.264	0.264	0.446	0.446	0.847	1.000	0.953	0.952	0.264	0.264
5	1	1	0.438	0.438	0.846	1.000	0.900	0.897	0.264	0.264	0.446	0.447	0.846	1.000	0.952	0.954	0.264	0.264
5	1	2	0.438	0.440	0.845	1.000	0.899	0.898	0.264	0.264	0.447	0.448	0.845	1.000	0.952	0.954	0.264	0.264
								$\vartheta_r 0 = 0.65$		$\delta_0 = 0.84$		$d_{r,i0} = 0.45$						
0.5	2	1	0.656	0.653	0.845	1.000	0.888	0.885	0.466	0.466	0.654	0.655	0.844	1.000	0.876	0.885	0.466	0.466
0.5	1	1	0.654	0.654	0.846	1.000	0.882	0.877	0.466	0.466	0.654	0.654	0.845	1.000	0.889	0.876	0.466	0.466
0.5	1	2	0.653	0.655	0.846	1.000	0.876	0.867	0.466	0.466	0.655	0.654	0.845	1.000	0.877	0.870	0.466	0.466
2	2	1	0.652	0.654	0.844	0.999	0.869	0.881	0.466	0.466	0.656	0.654	0.845	1.000	0.873	0.881	0.466	0.466
2	1	1	0.650	0.650	0.844	0.999	0.873	0.866	0.466	0.466	0.653	0.654	0.844	1.000	0.870	0.870	0.466	0.466
2	1	2	0.649	0.651	0.844	1.000	0.874	0.856	0.466	0.466	0.654	0.656	0.846	1.000	0.875	0.874	0.466	0.466
5	2	1	0.645	0.645	0.845	0.999	0.859	0.854	0.466	0.466	0.652	0.653	0.845	1.000	0.878	0.870	0.466	0.466
5	1	1	0.645	0.646	0.845	0.999	0.855	0.851	0.466	0.466	0.653	0.653	0.845	1.000	0.872	0.872	0.466	0.466
5	1	2	0.645	0.644	0.845	0.999	0.854	0.855	0.466	0.466	0.653	0.652	0.845	1.000	0.871	0.877	0.466	0.466

*Notes:*The averages of the memory estimated of the unobservable factors in both levels, the residual memory estimates, and the measure of consistency of the unobservable factors estimated are presented in the report. The DGP is the same as that of Table 2, except that now $i \in (20, 80)$ $T = 5000$. All experiments are based on 1000 replications.

Conclusions from Tables 2-5 do not change by increasing N_r and reducing T. The latter configuration can be more related with some macroeconomic applications. In this light, we simulate one replication of the model 1 with $R = 2$, $N_r = 300$, $T = 150$, with the medium signal-to-noise ratio and $\vartheta_{r0} = 0.6$, $\delta_0 = 1$, and $d_{r,i0} = 0.25$. Following Wang (2010), once we get $\hat{G}$,$\hat{F}$,$\hat{\mu}$, and $\hat{\Lambda}$, we project the true factors on the estimated factors to find the rotation matrix, $\hat{Q}_G = \left(\hat{G}'\hat{G}\right)^{-1}\hat{G}'G$. Then we use $\left(\hat{Q}_G\right)^{-1}$ to rotate factor loadings. Figure 2 displays the precision of the projected estimators with the true ones.

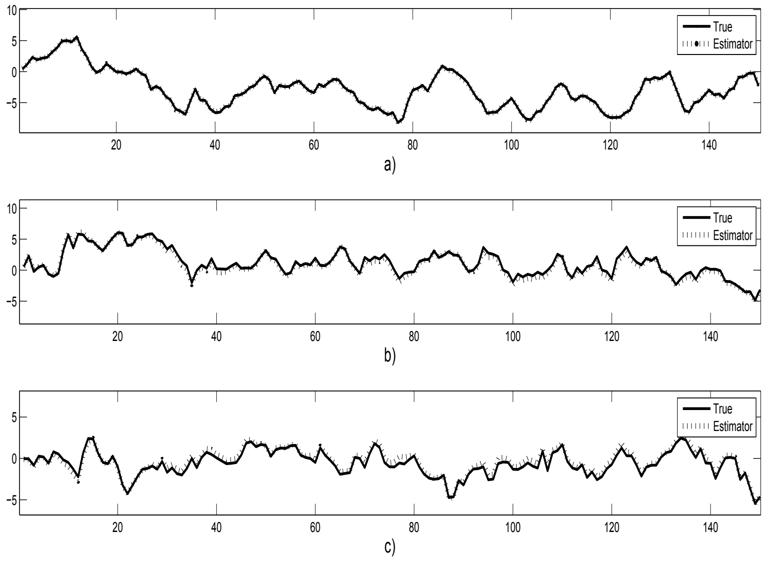

Figure 2: The dashed lines are the estimators for factors projected onto the true ones (solid lines). Global factor in panel a), Regional factors are in panel b) and c).

The third simulation study, Table 6, presents two simple Monte Carlo experiments to show two specific points to be taken into account in our methodology. First, when the memory of the residuals $d_{r,i}$ is in the nonstationary region, the accuracy of the model in (1) to identify the regional factor decreases considerably when $d_{r,i}$ increases. However, it is still possible to extract the global factor as well as its memory level and the memory levels of the residuals reasonably well[3]. Second, when there is no fractional cointegration, neither the factors nor the mem-

[3]Note that the initial values of the global factor are based on CCA but those of regional factors on PCA. In this sense, our findings seem to indicate that CCA may play a key role here in comparison to PCA, although this claim requires further justification and is left for future research.

Table 5: Second Monte Carlo simulation with $T \in (250, 1000, 5000)$, $N_r = 20$, and $R = 4$.

σ	ϑ_{10}	ϑ_{20}	ϑ_{30}	ϑ_{40}	δ_0	$d_{r,i0}$	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\vartheta}_3$	$\hat{\vartheta}_4$	$\hat{\delta}$	R^2_g	R^2_{R1}	R^2_{R2}	R^2_{R3}	R^2_{R4}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$	$\bar{\hat{d}}_{R3}$	$\bar{\hat{d}}_{R4}$
													T=250							
0.5	0.3	0.4	0.5	0.6	0.6	0	0.28	0.37	0.47	0.57	0.59	0.99	0.92	0.92	0.92	0.92	0.06	0.06	0.06	0.06
1	0.3	0.4	0.5	0.6	0.6	0	0.29	0.39	0.48	0.58	0.60	1.00	0.97	0.96	0.96	0.94	0.06	0.06	0.06	0.06
2	0.3	0.4	0.5	0.6	0.6	0	0.29	0.39	0.48	0.58	0.60	1.00	0.98	0.98	0.96	0.95	0.06	0.06	0.06	0.06
0.5	0.5	0.6	0.7	0.8	1	0.25	0.48	0.58	0.68	0.79	0.98	1.00	0.91	0.90	0.86	0.83	0.25	0.25	0.25	0.25
1	0.5	0.6	0.7	0.8	1	0.25	0.50	0.59	0.69	0.80	0.98	1.00	0.94	0.92	0.89	0.84	0.25	0.25	0.25	0.25
2	0.5	0.6	0.7	0.8	1	0.25	0.50	0.60	0.70	0.79	0.98	1.00	0.96	0.93	0.88	0.84	0.25	0.25	0.25	0.25
													T=1000							
0.5	0.2	0.3	0.4	0.5	0.5	0	0.20	0.30	0.40	0.50	0.50	0.99	0.92	0.94	0.95	0.94	0.06	0.06	0.06	0.06
1	0.2	0.3	0.4	0.5	0.5	0	0.20	0.30	0.40	0.50	0.51	1.00	0.98	0.98	0.98	0.96	0.06	0.06	0.06	0.06
2	0.2	0.3	0.4	0.5	0.5	0	0.20	0.30	0.41	0.50	0.51	1.00	0.99	0.99	0.98	0.97	0.06	0.06	0.06	0.06
0.5	0.4	0.5	0.6	0.7	0.9	0.25	0.40	0.50	0.60	0.71	0.91	1.00	0.93	0.92	0.90	0.85	0.26	0.26	0.26	0.26
1	0.4	0.5	0.6	0.7	0.9	0.25	0.40	0.51	0.61	0.71	0.91	1.00	0.96	0.95	0.92	0.87	0.26	0.26	0.26	0.26
2	0.4	0.5	0.6	0.7	0.9	0.25	0.40	0.51	0.61	0.71	0.91	1.00	0.98	0.95	0.92	0.86	0.26	0.26	0.26	0.26
													T=5000							
0.5	0.15	0.25	0.35	0.45	0.45	0	0.14	0.24	0.34	0.44	0.45	0.99	0.93	0.94	0.95	0.96	0.06	0.06	0.06	0.06
1	0.15	0.25	0.35	0.45	0.45	0	0.15	0.24	0.34	0.44	0.45	1.00	0.98	0.98	0.98	0.98	0.06	0.06	0.06	0.06
2	0.15	0.25	0.35	0.45	0.45	0	0.15	0.25	0.35	0.44	0.45	1.00	0.99	0.99	0.99	0.98	0.06	0.06	0.06	0.06
0.5	0.35	0.45	0.55	0.65	0.85	0.25	0.34	0.44	0.55	0.65	0.86	1.00	0.94	0.94	0.92	0.88	0.26	0.26	0.26	0.26
1	0.35	0.45	0.55	0.65	0.85	0.25	0.35	0.45	0.55	0.66	0.86	1.00	0.98	0.96	0.93	0.88	0.26	0.26	0.26	0.26
2	0.35	0.45	0.55	0.65	0.85	0.25	0.35	0.45	0.55	0.65	0.86	1.00	0.99	0.97	0.93	0.88	0.26	0.26	0.26	0.26

*Notes:*The averages of the memory estimated of the unobservable factors in both levels, the residual memory estimates, and the measure of consistency of the unobservable factors estimated are presented in the report. The DGP is the same as that of Table 2, except that now $r = 4$, $i = 20$, and $T \in (250, 1000, 5000)$. All experiments are based on 1000 replications.

Table 6: Third Monte Carlo simulation with $T = 5000$, $N_r = 20$, and $R = 2$.

ϑ_r	δ	d	$\hat{\vartheta}_1$	$\hat{\vartheta}_2$	$\hat{\delta}$	R^2_G	R^2_{R1}	R^2_{R2}	$\bar{\hat{d}}_{R1}$	$\bar{\hat{d}}_{R2}$
Experiment: Nonstationary d_{ri}										
0.7	0.8	0.6	0.75	0.75	0.86	0.99	0.84	0.82	0.67	0.67
1	1.2	0.8	1.05	1.05	1.07	0.99	0.61	0.51	0.87	0.87
1.3	1.5	1	1.15	1.16	1.02	1.00	0.42	0.40	1.07	1.07
1.5	1.5	1	1.13	1.10	1.06	1.00	0.45	0.51	1.06	1.06
Experiment: Non-fractional cointegration										
0.3	0.5	0.6	0.39	0.38	0.55	0.97	0.83	0.85	0.68	0.68
0.5	0.7	1	0.80	0.81	0.85	0.42	0.11	0.08	1.03	1.03

*Notes:*The averages of the memory estimated of the unobservable factors in both levels, the residual memory estimates, and the measure of consistency of the unobservable factors estimated are presented in the report. The DGP is the same as that of Table 2. All experiments are based on 1000 replications.

ory level of the global and regional factors can be precisely estimated but we can consistently estimate the memory of the residuals. Since we are able to estimate the memory of residuals even in the absence of fractional cointegration, we can still analyze whether $\hat{d}_{r,i} > max\left(\hat{\vartheta}_r, \hat{\delta}\right)$ in order to be sure that global and regional factors are accurately estimated.

4.2 The number of regional and global factors

We now present a Monte Carlo experiment to show the reliability of the methodology proposed to estimate the number of regional and global factors in relation to the presentation in Section 3. We design our simulation study using the same framework as before.

Tables 7-9 show that in general, the information criteria IC1, IC2, and IC3 perform better than PC1, PC2, and PC3. Furthermore, the number of factors using IC1, IC2, and IC3 is always consistently estimated when variables are fractionally differenced by $\varsigma = max\left(\delta_0, \vartheta_{r,i0}\right)$. Only in cases when $d_{r,i0} \leq 1$, the number of factor is accurately estimated taking the first differences of the variables. The original variables can be used only in the specific case when $d_{r,i0} = 0$ although the performance of the number of factors does not diminish considerably in cases when $d_{r,i0} < 0.5$. Since Tables 7 and 8 consider two regions, we have three different blocks of data, $B_{R_1 \cup R_2}$, B_{R_1}, and B_{R_2} as explained before. Table 7 includes the case of only one global factor and one regional factor in each region, consequently the actual number of static factors in each block are $s_{B_{R_1 \cup R_2}} = 3$, $s_{B_{R_1}} = 2$, and $s_{B_{R_2}} = 2$ as represented in Figure 1. Table 8 considers the case of two global factors and two regional factors, then $s_{B_{R_1 \cup R_2}} = 6$, $s_{B_{R_1}} = 4$, and $s_{B_{R_2}} = 4$. When considering three regions in Table 9 with one global factor and one regional factor in each region, we have seven different blocks with the number of static factors as follows $s_{B_{R_1 \cup R_2 \cup R_3}} = 4$, $s_{B_{R_1}} = 2$, $s_{B_{R_2}} = 2$, $s_{B_{R_3}} = 2$, $s_{B_{R_1 \cup R_2}} = 3$, $s_{B_{R_1 \cup R_2}} = 3$, and $s_{B_{R_2 \cup R_3}} = 3$. Finally, the number of global and the regional factors can be obtained by the inclusion-exclusion principle.

5 An Application to Nord Pool Power Market

In this section, we provide an application of our methodology to study price co-movements in the Nord Pool power spot market.

Over the past few decades, a liberalization of power markets has emerged. Power companies produce electricity power from many different sources (hydro, thermal, nuclear, wind, and solar systems) in order to provide competitive prices and ensure production efficiency. From an economic perspective, electricity markets seek to match the supply and demand in order to find a market clearing price. Moreover, spot prices exhibit seasonality at daily and weekly levels by daily activities either on working or non-working days, and at a yearly level due to changing weather conditions throughout the year. Such prices also present irregular cyclical factors which are associated with cyclical movements in the economy or long-term climate trends whereas several spikes are caused by some anticipated special dates (Christmas, national holidays, etc.) and unanticipated days intrinsically originated in the market. Weron (2007) reports these stylized facts as well as an overview of statistical methods used in the literature.

Another feature that has received considerable attention is the presence of a hyperbolic decay of the autocovariances of electricity prices. In this light, Haldrup and Nielsen (2006) use Phillips-Perron and KPSS tests to suggest that neither an I(0) nor I(1) process is appropriate for electricity prices. They point out that Nord Pool prices are characterized by a high degree of long memory. Along this line, Koopman et al. (2007) consider general seasonal periodic regressions with ARFIMA-GARCH disturbances to analyze daily spot prices.

Table 7: Number of Common Factors. 2 regions. 3 blocks of data. (N = 40, T = 500 and k_{max} = 10). 1 Global factor and 1 regional factor in each region.

	Neglected memory			First Difference			Fractional Differencing using δ_0		
	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$
	$d_{r,i0} = 1.5$, $\delta_0 = 2$, and $\vartheta_{r0} = 1.8$.								
IC1	10.00	10.00	10.00	6.71	5.54	5.51	3.00	2.00	2.00
IC2	10.00	10.00	10.00	6.62	5.25	5.32	3.00	2.00	2.00
IC3	10.00	10.00	10.00	7.44	6.21	6.23	3.00	2.00	2.00
PC1	10.00	10.00	10.00	8.41	9.71	9.81	3.00	6.41	6.41
PC2	10.00	10.00	10.00	8.36	9.72	9.72	3.00	6.11	6.21
PC3	10.00	10.00	10.00	8.83	9.91	9.93	3.00	6.92	6.91
	$d_{r,i0} = 0.4$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.7$.								
IC1	3.31	2.25	2.23	3.00	2.00	2.00	3.00	2.00	2.00
IC2	3.27	2.23	2.21	3.00	2.00	2.00	3.00	2.00	2.00
IC3	3.43	2.29	2.26	3.00	2.00	2.00	3.00	2.00	2.00
PC1	4.65	7.47	7.45	3.00	4.40	4.40	3.00	5.48	5.49
PC2	4.53	7.27	7.28	3.00	4.11	4.10	3.00	5.21	5.20
PC3	4.97	7.84	7.84	3.00	4.95	4.96	3.00	6.06	6.03
	$d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.								
IC1	6.75	5.51	5.57	3.00	2.00	2.00	3.00	2.00	2.00
IC2	6.56	5.22	5.27	3.00	2.00	2.00	3.00	2.00	2.00
IC3	7.37	6.20	6.15	3.00	2.00	2.00	3.00	2.00	2.00
PC1	8.42	9.77	9.73	3.00	3.96	3.94	3.00	5.35	5.35
PC2	8.28	9.69	9.64	3.00	3.70	3.67	3.00	5.07	5.10
PC3	8.85	9.89	9.87	3.00	4.54	4.53	3.00	5.91	5.90

Notes: The DGP is the same as that of Table 2. $s_{B_{R_1 \cup R_2}}$, $s_{B_{R_1}}$, and $s_{B_{R_2}}$ are the averages of the number of factors estimated in each block by the inclusion-exclusion principle. We compare three cases: i) when memory is neglected, ii) taking the first difference on $y_{r,it}$, and iii) fractional differencing $y_{r,it}$ with $\delta_0 = max(\delta_0, \vartheta_{r0})$. IC1, IC2, IC3, PC1, PC2, and PC3 are the information criteria of Bai and Ng (2002). All experiments are based on 1000 replications.

Table 8: Number of Common Factors. 2 regions. 3 blocks of data. (N = 40, T = 500 and k_{max} = 10). 2 Global factors and 2 regional factors in each region.

	Neglected memory			First Difference			Fractional Differencing using δ_0		
	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$	$s_{B_{R_1 \cup R_2}}$	$s_{B_{R_1}}$	$s_{B_{R_2}}$
	$d_{r,i0} = 1.5$, $\delta_0 = 2$, and $\vartheta_{r0} = 1.8$.								
IC1	10.00	10.00	10.00	6.78	4.83	4.80	6.00	4.00	4.00
IC2	10.00	10.00	10.00	6.68	4.77	4.75	6.00	4.00	4.00
IC3	10.00	10.00	10.00	7.06	4.99	4.98	6.00	4.00	4.00
PC1	10.00	10.00	10.00	7.84	9.12	9.11	6.00	6.95	6.99
PC2	10.00	10.00	10.00	7.73	8.99	8.98	6.00	6.74	6.75
PC3	10.00	10.00	10.00	8.15	9.40	9.39	6.00	7.40	7.45
	$d_{r,i0} = 0.4$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.7$.								
IC1	6.05	4.08	4.08	6.00	4.00	4.00	6.00	4.00	4.00
IC2	6.04	4.07	4.07	6.00	4.00	4.00	6.00	4.00	4.00
IC3	6.08	4.10	4.10	6.00	4.00	4.00	6.00	4.00	4.00
PC1	6.47	7.90	7.92	6.00	5.21	5.22	6.00	6.13	6.12
PC2	6.40	7.73	7.73	6.00	4.98	5.00	6.00	5.89	5.87
PC3	6.66	8.22	8.21	6.00	5.68	5.68	6.00	6.60	6.61
	$d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.								
IC1	8.77	7.74	7.73	6.00	4.00	4.00	6.00	4.00	4.00
IC2	8.59	7.39	7.37	6.00	4.00	4.00	6.00	4.00	4.00
IC3	9.24	8.41	8.53	6.00	4.00	4.00	6.00	4.00	4.00
PC1	9.38	9.89	9.89	6.00	4.88	4.90	6.00	5.98	6.01
PC2	9.27	9.84	9.85	6.00	4.67	4.69	6.00	5.74	5.79
PC3	9.68	9.95	9.96	6.00	5.33	5.35	6.00	6.45	6.49

Notes: The DGP is the same as that of Table 7. All experiments are based on 1000 replications.

Table 9: Number of Common Factors. 3 regions. 7 blocks of data. (N = 40, T = 500 and k_{max} = 10). 1 Global factor and 1 regional factor in each region. Estimation is performed in first differences. $d_{r,i0} = 0.6$, $\delta_0 = 1$, and $\vartheta_{r0} = 0.8$.

	$^{s}B_{R_1 \cup R_2 \cup R_3}$	$^{s}B_{R_1}$	$^{s}B_{R_2}$	$^{s}B_{R_3}$	$^{s}B_{R_1 \cup R_2}$	$^{s}B_{R_1 \cup R_2}$	$^{s}B_{R_2 \cup R_3}$
IC1	4.00	2.00	2.00	2.00	3.00	3.00	3.00
IC2	4.00	2.00	2.00	2.00	3.00	3.00	3.00
IC3	4.00	2.00	2.00	2.00	3.00	3.00	3.00
PC1	4.00	5.38	5.35	5.32	3.00	3.00	3.00
PC2	4.00	5.09	5.09	5.07	3.00	3.00	3.00
PC3	4.00	5.92	5.88	5.90	3.00	3.00	3.00

Notes: The DGP is the same as that of Table 7. All experiments are based on 1000 replications.

Although daily average prices are widely studied in the literature due to the role played in the so-called day-ahead market, it would also be of interest to disaggregate electricity prices in order to strengthen the respective prediction as Ramanathan et al. (1997) stress. In this regard, Raviv et al. (2015) also point out that the daily average of the disaggregate hourly forecasts contain useful information to study the daily average price in the Nord Pool market. In addition, it is habitually overlooked when modeling the hourly prices that the vector of 24 hourly prices is determined simultaneously in the day-ahead market. The latter means that a proper form of the data set would be a panel of prices with a natural ordering in the cross section dimension instead of a single time series since consecutive prices are determined simultaneously.

Examining in detail the hourly electricity prices implies the study of a complex dependence structure in the market, which has not been extensively considered in the literature. A natural way to take into account such dependence is with a Vector Autoregressive (VAR) approach, however, that would lead to the so-called 'curse of dimensionality' problem and hence it is also of interest to reduce dimensionality. Panel data and factor models are standard tools to analyze high dimensional data and have been recently used in electricity markets (see e.g. Alonso et al. (2011), Dordonnat et al. (2012) and Raviv et al. (2015)).

The 'Panel Analysis of Nonstationarity in Idiosyncratic and Common components' (PANIC) is an alternative way to study the complexity of electricity prices. In this regard, Ergemen et al. (2015) very recently model the complex dynamics of Nord Pool electricity prices in the Elspot market by considering the models proposed by Ergemen and Velasco (2017) and Ergemen (2016) which allow for fractionally integrated panels with fixed effects and cross-section dependence. We may say in a sense that the application of this paper is in line with one of the findings in Ergemen et al. (2015) that suggests a fractional cointegrating relationship in the panel of electricity prices and their main unobservable common factor although they do not consider an energy market divided by some regions but instead work with a reference price for the whole energy system, i.e. the system price.

A possible limitation of the aforementioned study is the use of these system prices which are the unconstrained equilibrium price for the entire Nordic region disregarding the available transmission capacity between the bidding areas. However, the Elspot market is divided into several bidding areas due to system prices not clearing all regions within the Nordic market.

Another possible drawback when analyzing the entire Nordic market is that empirical studies, which include factor models in their analysis, assume that the common factors affect all regions of the system without taking into account some region-specific characteristics. In principle it is natural to extract common factors of each specific bidding area of the Nord Pool market and analyze them separately. However it is not clear whether there exist global common factors affecting all areas and provoke severe loss of efficiency when trying to identify common factors. Hence our interest in studying the hourly prices dynamics by considering each bidding area of the Nord Pool market.

In the present paper, the data set under consideration are $R = 12$ balanced panels consisting of $N_r = 24$ hourly prices for each day for the period January 1, 2012, to December 31, 2014, and thus yielding a total of $T = 1096$ daily observations in each panel. We consider 12 panels since we analyze 12 bidding areas: five Norwegian bidding areas (NO1-NO5), Western Denmark (DK1), Eastern Denmark (DK2), four Swedish bidding areas (SE1-SE4), and Finland (FI). All bidding areas are connected. The series are downloaded from the Nord Pool ftp server. The prices are denominated in euros per Mwh of load. Following Ergemen et al. (2015), the series are prefiltered by

$$y_{it} = \alpha_{i0} + \alpha_{i1}\ t + \alpha_{i2}\ D_t + \mathbb{B}_t' A_i + \alpha_{i3}\ \cos(\frac{2\pi t}{365}) + \alpha_{i4}\ \cos(\frac{2\pi t}{7}) + \alpha_{i5}\ \cos(\frac{2\pi t}{3.5}) + y_{it}^*, \quad (6)$$

where $\mathbb{B}_t$ is a vector of shift dummies, which captures level changes caused by structural breaks. D_t is a dummy variable for holidays that takes the value of 1 if any of the countries participating in the Nord Pool system suspends or reduces normal business activities by custom or law, and 0 otherwise. The data for non-working days in each of the countries of the Nord Pool System is extracted from Bloomberg, which is then incorporated into the analysis due to the strong effect of holidays in the electricity market, see Ergemen et al. (2015) for more details.

As explained in Section 3, the number of different blocks from the regional data considerably increases with the number of regions. In our case, since the number of bidding areas to be analyzed are 12, we would have $2^{12} - 1 = 4,095$ different blocks. To avoid such complexity, we take advantage of the correlation showed by the daily regional prices to establish only four regions as follows: Region 1 = (DK1, DK2), Region 2 = (NO1, NO2, NO5), Region 3 = (NO3, NO4, SE1, SE2, SE3, SE4), and Region 4 = FI. Table 10 shows the correlation matrix of the daily prices whereas Figure 3 displays the map of the Nord Pool market with these four regions. Note that each region consists of neighboring bidding areas. The bidding area corresponding to Finland presents much more spikes than any other area and may decrease such correlations.

Table 10: Correlation matrix of the daily prices in each bidding area of Nord Pool power market.

	DK1	DK2	NO1	NO2	NO3	NO4	NO5	SE1	SE2	SE3	SE4	FI
DK1	1.00											
DK2	0.93	1.00										
NO1	0.57	0.61	1.00									
NO2	0.56	0.57	0.98	1.00								
NO3	0.68	0.73	0.87	0.84	1.00							
NO4	0.66	0.72	0.88	0.85	0.99	1.00						
NO5	0.54	0.57	0.99	0.99	0.85	0.85	1.00					
SE1	0.72	0.77	0.84	0.81	0.98	0.97	0.82	1.00				
SE2	0.72	0.77	0.84	0.81	0.98	0.97	0.82	1.00	1.00			
SE3	0.74	0.79	0.83	0.80	0.96	0.96	0.80	0.99	0.99	1.00		
SE4	0.79	0.86	0.79	0.75	0.91	0.90	0.76	0.93	0.93	0.95	1.00	
FI	0.66	0.70	0.66	0.63	0.81	0.80	0.63	0.83	0.83	0.85	0.78	1.00

For each of the four regions we compute the hourly regional prices by taking the simple average of hourly prices corresponding to the bidding areas that define the region. Naturally, we still have 24 hourly prices. It is possible to consider a weighted average of the prices by considering the available transmission capacity but we work with the simplest average to focus on the main ideas.

We estimate the memory, $\varepsilon_{r,i}$, of each one of the hourly regional prices with the Extended Local Whittle procedure. Each hourly regional price is fractionally dif-

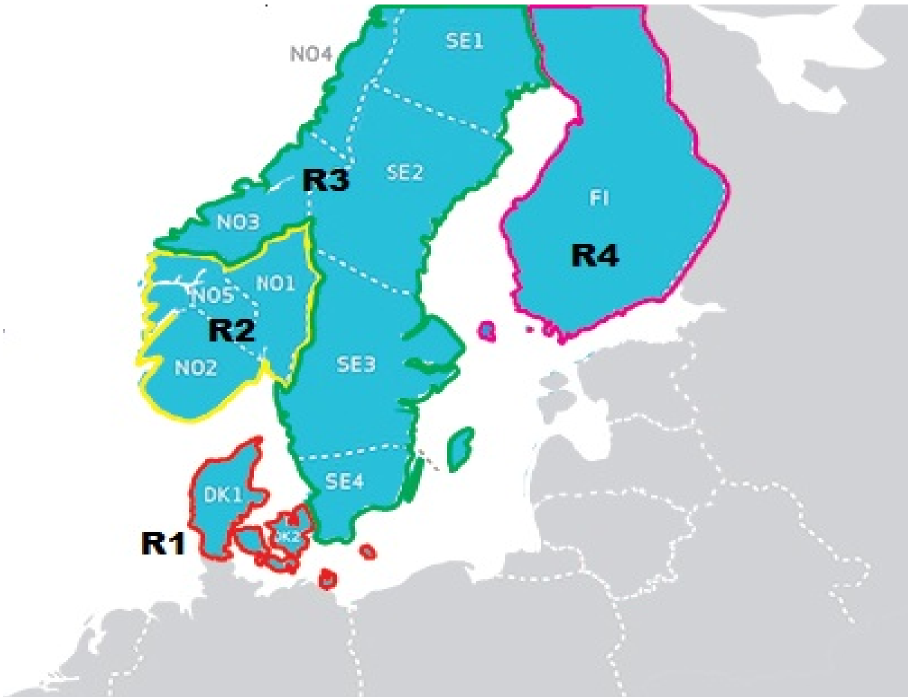

Figure 3: Nord Pool power market divided by the new regions.

ferenced by its respective estimated memory, $\hat{\varepsilon}_{r,i}$, so that the number of global and regional static factors can be estimated as described in Section 3.

To estimate the number of regional and global factors, we use the procedure proposed by Alessi et al. (2010). This procedure improves the penalization in the criteria IC1 and IC2 of Bai and Ng (2002) introducing a tuning multiplicative constant in the penalty function under the same set of assumptions that lead to heteroskedasticity-robust inference. Theorem 2 in Bai and Ng (2002) is still valid and consequently our methodology can also be applied with the information criteria of Alessi et al. (2010).

Table 11 presents the number of factors estimated in each one of the 15 blocks. Furthermore, using the inclusion-exclusion principle, we get the number of global and regional factors. Note that for computing the number of regional static factors, we need to compute the number of static factors in each one of the blocks of regions. Edward's diagram in Figure 4 displays how the static factors are accommodated in each block. Particularly, such a diagram shows that we find one global factor and two regional factors in each one of the regions. Note that the sum of all the static factors identified in the Edward's diagram corresponds to the number of static factors estimated in the quadruple-wise block $B_{R_1 \cup R_2 \cup R_3 \cup R_4}$.

Table 11: Number of static factors in the 15 blocks formed.

Individual blocks		Pairwise blocks		Triple-wise blocks		Quadruple-wise block	
s_{R_1}	7	$s_{B_{R_1} \cup B_{R_2}}$	11	$s_{B_{R_1} \cup B_{R_2} \cup B_{R_3}}$	14	$s_{B_{R_1} \cup B_{R_2} \cup B_{R_3} \cup B_{R_4}}$	16
s_{R_2}	7	$s_{B_{R_1} \cup B_{R_3}}$	12	$s_{B_{R_1} \cup B_{R_3} \cup B_{R_4}}$	14		
s_{R_3}	7	$s_{B_{R_1} \cup B_{R_4}}$	11	$s_{B_{R_1} \cup B_{R_2} \cup B_{R_4}}$	14		
s_{R_4}	7	$s_{B_{R_2} \cup B_{R_3}}$	11	$s_{B_{R_2} \cup B_{R_3} \cup B_{R_4}}$	14		
		$s_{B_{R_2} \cup B_{R_4}}$	11				
		$s_{B_{R_3} \cup B_{R_4}}$	11				

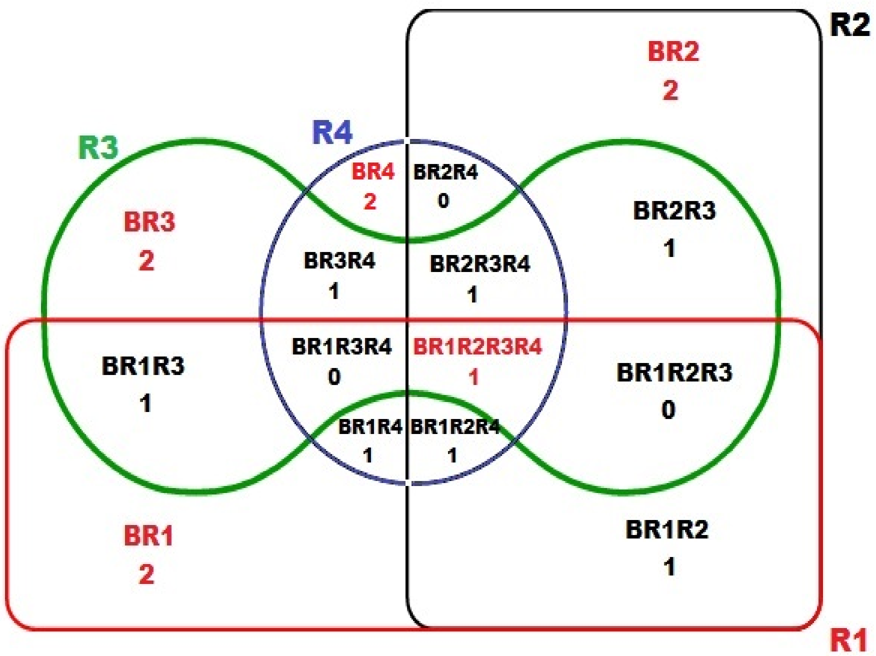

Figure 4: Number of factors in the 15 blocks.

Now we estimate the model specified in (1) by the methodology proposed in Section 2.2. We fix the number of global factors to be one and the regional factors to be two in all four regions. Figure 5 shows the global factor and its loadings for each region. The first panel in Figure 5 also displays the filtered system daily prices by the same filtering model (6) as before.

As seen from Figure 5, the global factor tends to be highly persistent. The global factor loadings show a regular behavior among bidding areas. Loadings are positive and larger overnight indicating that the global factor plays a key role from 12 a.m.-7 a.m. and from 10 p.m.-12 p.m. Levels of the loadings are similar across regions. Figure 5 shows that the global factor fits well to the filtered system prices. Furthermore, the correlation between the global factor and the filtered system price is around 0.85. The estimated memory of the filtered system prices is 0.77 whereas the estimated memory of the global factor is 0.82. Our findings indicate that the global factor may be interpreted as the system price even when we have reduced from 12 to only 4 bidding regions.

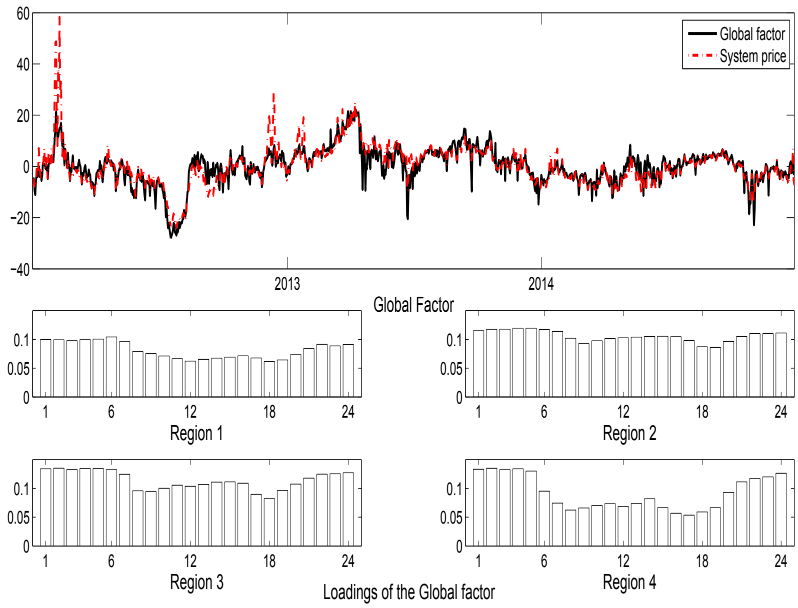

Figure 5: Global common component of Nord Pool bidding areas.

Figures 6 and 7 present the regional loadings and both regional factors of each region. Figure 6 indicates that the regional factors play a key role during working hours explaining much more of the variability, mainly the second regional factors. Consequently, for each hourly price, the commonality of the regional factor implies a small subtraction over the commonality of this hour considering only the global factor. On the other hand, during working hours, the commonality of regional factors explains more of the variability than the global factor which only represents a small correction. Furthermore, it is apparent that regional factors also seem to be highly persistent. Table 12 shows the correlation among first regional factor as well as second regional factors. All correlations are low indicating that regional factors differ region by region.

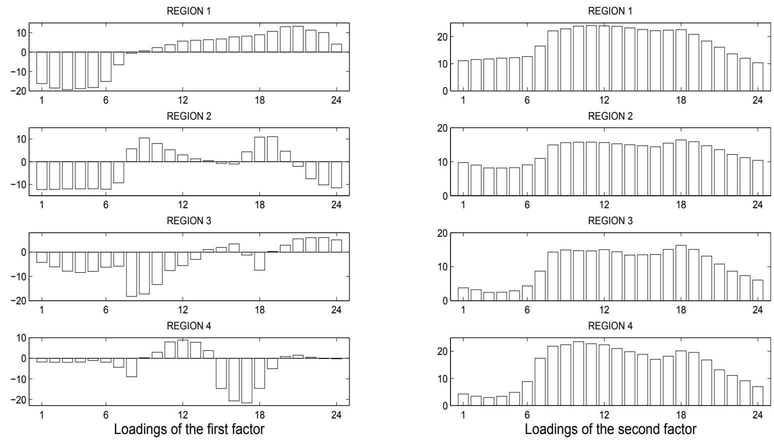

Figure 6: Loadings of the regional factors.

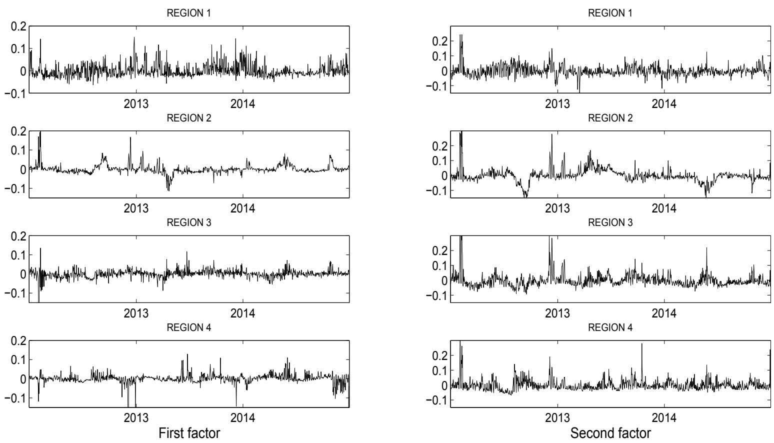

Figure 7: Regional factors.

Table 12: Correlation among regional factors.

	First regional factors			
	Region 1	Region 2	Region 3	Region 4
Region 1	1			
Region 2	0.13	1		
Region 3	0.35	0.01	1	
Region 4	0.01	0.06	0.02	1
	Second regional factors			
	Region 1	Region 2	Region 3	Region 4
Region 1	1			
Region 2	0.33	1		
Region 3	0.63	0.60	1	
Region 4	0.50	0.32	0.66	1

We study whether a fractional cointegration relationship exists in our analysis, which ensures that the memory of the residuals stay in the stationary region to verify that our model is consistently estimated. We collect the global and regional factors $(\hat{G}_t, \hat{F_{r,t}})$ and estimate the fractional memory parameters with the Extended Local Whittle procedure proposed by Abadir et al. (2007). Figure 8 displays that the global factor is more persistent than regional factors. Regional factors of Region 2 are more persistent than the other regions while the Danish region, Region 1, shows less persistence in both regional factors.

Fractional cointegration relationship is confirmed given that for each region and each hour of the day $\hat{d}_{r,i} \leq \hat{\delta}$. Figure 9 shows that persistence levels of the residuals of model in (1) have decreased once we have taken into account the strong dependence of the hourly electricity prices analyzed with the global and regional factors estimated.

For practical purposes, let us assume that an analyst is interested in studying the block of panels in our application. Practitioners would estimate a number of common factors but neglecting a multi-level structure as we proposed in this paper. As we have already discussed before, that common factor could be mixed up with the regional ones provoking that the estimation of the actual pervasive factor may be hampered. For comparison, we estimate a common factor but neglecting a multi-level structure. We estimate a common factor by principal components after taking the first differences in all the 24 prices of the 4 regions. Then, we integrate back the factor estimates to obtain the original estimates. Figure 10 displays such

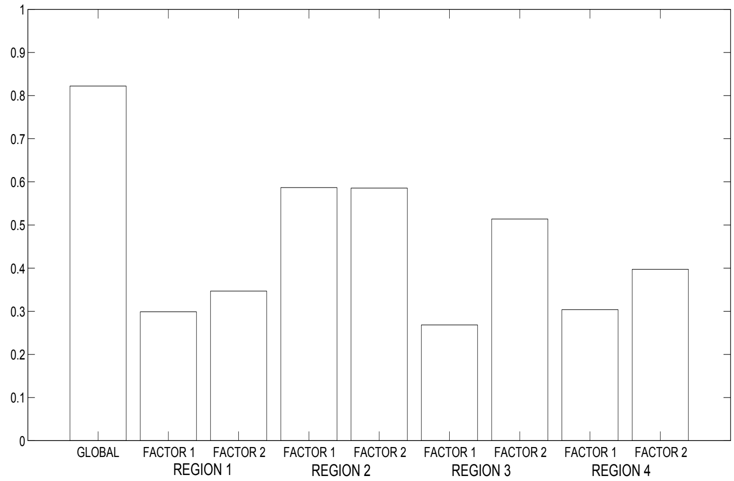

Figure 8: Memory estimates of the global and regional factors. The number of Fourier frequencies used is m = $T^{0.7}$ with T = 1096 corresponding to m = 134. The standard error of the univariate estimates is 0.043.

a comparison.

To conclude, the number of factors in each one of the blocks represented in Figure 4 can be used to extend the model in (1). For instance, consider the case of the three first bidding regions of the Nord Pool power market, R1, R2, and R3. We may extend our model as

$$\begin{pmatrix} y_{R_1,\cdot t} \\ y_{R_2,\cdot t} \\ y_{R_3,\cdot t} \end{pmatrix} = \begin{pmatrix} \Gamma_1 & \Lambda_1 & 0 & 0 & \kappa_{1,12} & 0 & \kappa_{1,13} \\ \Gamma_2 & 0 & \Lambda_2 & 0 & \kappa_{2,12} & \kappa_{2,23} & 0 \\ \Gamma_3 & 0 & 0 & \Lambda_3 & 0 & \kappa_{3,23} & \kappa_{3,13} \end{pmatrix} \begin{pmatrix} G_{123,t} \\ F_{1,t} \\ F_{2,t} \\ F_{3,t} \\ F_{12,t} \\ F_{23,t} \\ F_{13,t} \end{pmatrix} + \begin{pmatrix} u_{1,\cdot t} \\ u_{2,\cdot t} \\ u_{3,\cdot t} \end{pmatrix},$$

where we would add to model specified in (1), the common factors, $F_{12,t}$, $F_{23,t}$, and $F_{13,t}$, corresponding to blocks $B_{R_1 \cap R_2}$, $B_{R_2 \cap R_3}$, and $B_{R_1 \cap R_3}$, respectively. The third block in the loading matrix, κ's, would be the respective blocks' loading factors. In future research, we plan to focus on estimating this kind of extended models in order to analyze in depth the interaction among blocks of regions in a multi-level factor model. In principle, the methodology proposed in this paper can be used to estimate the new model after incorporating more steps in the procedure, however it would be also necessary to add more restrictions and assumptions in

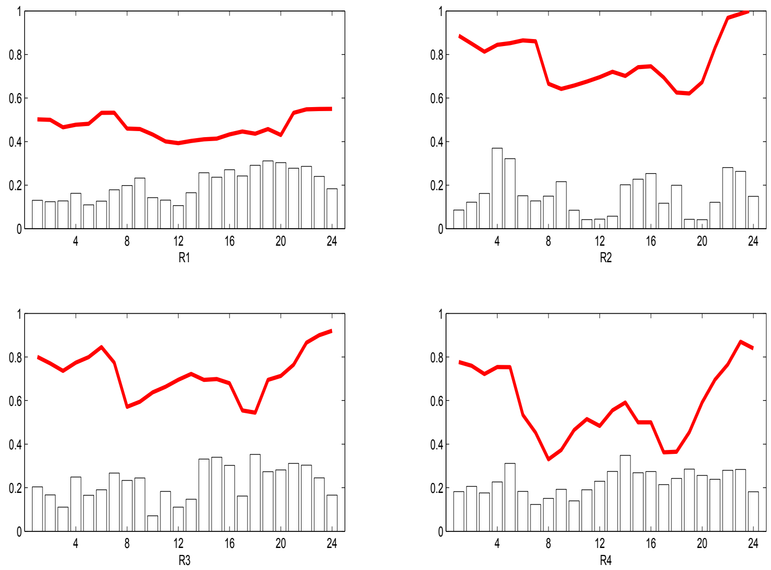

Figure 9: Comparative of the residual integration order estimates (bars) with the memory estimates of the filtered regional prices for each region (red line).

order to identify the model.

6 Concluding remarks

In this paper we have considered a dynamic multi-level factor model, which allows for both pervasive and nonpervasive common factors. The multilevel factor structure and model innovations are allowed to exhibit short-memory dynamics and long-range dependence without restrictions on them being either stationary I(0) or nonstationary I(1) processes. In order to estimate the model, we have proposed a parsimonious two-step procedure and discussed how the number of global and regional factors can be estimated. Through an extensive simulation study, we have shown that the methodology performs well in small samples and we then applied to the Nord Pool electricity market. While the model in (1) is quite general in that there is allowance for both multilevel factor structure and short-term as well as long-range dynamics, it can nevertheless be extended to account for parametric spatial dependence that would be useful in the analysis economic unions and spillover effects. Furthermore, forecasting studies can be undertaken using (1) to obtain more specific information exploiting the differences between global and re-

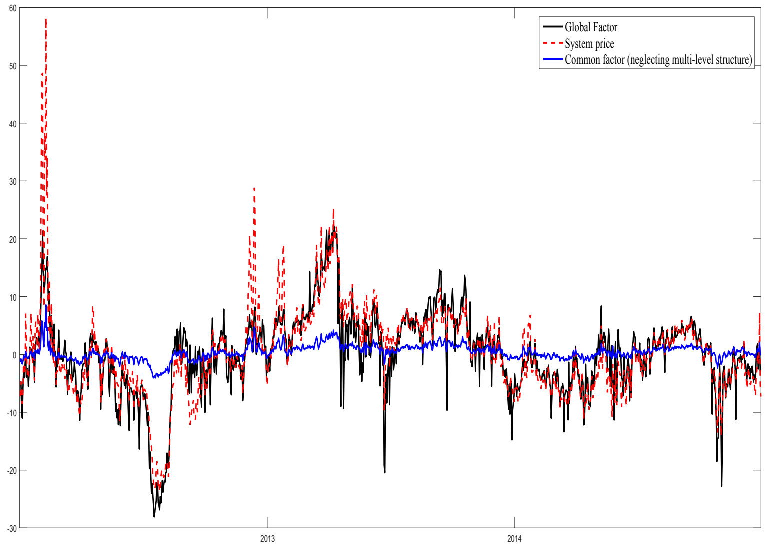

Figure 10: Comparison of a global factor and a common factor neglecting a multi-level structure.

gional effects. These ideas loom largely on this already lengthy paper and are left for future research.

7 Appendix

7.1 Proof of Theorem 2.1

Taking the cross-sectional and regional average of the model in (1),

$$\bar{\bar{y}}_t = \bar{\bar{\mu}}' G_t + \bar{\bar{\lambda}}' F_t + \bar{\bar{\epsilon}}_t$$

the quantities $\bar{\bar{\mu}}$ and $\bar{\bar{\lambda}}$ do not depend on i or r so $\bar{\bar{y}}_t$ is a pure time series that is integrated of order θ. Therefore the result can be shown following exactly the same steps as in Shimotsu (2010) under Assumptions A-E. □

7.2 Proof of Theorem 2.2

First, let us write

$$\begin{aligned}
\hat{F}_t^*(\hat{\theta}) - F_t^*(\theta) H_{F^*} &= \hat{F}_t^*(\hat{\theta}) - \hat{F}_t^*(\theta) + \hat{F}_t^*(\theta) - F_t^*(\theta) H_{F^*} \\
&= \hat{F}_t^*(\hat{\theta}) - \hat{F}_t^*(\theta) + O_p\left(\frac{1}{\min\left\{\sqrt{N}, T\right\}}\right)
\end{aligned}$$

where the $O_p(\cdot)$ term on the RHS appears because for fixed t,

$$\min\left\{\sqrt{N}, T\right\}\left(\hat{F}_t^*(\theta) - F_t^*(\theta) H_{F^*}\right) = O_p(1)$$

following Lemma 1(b) of Bai and Ng (2004).

Furthermore, applying the Mean Value Theorem,

$$\begin{aligned}
\hat{F}_t^*(\hat{\theta}) - \hat{F}_t^*(\theta) &= \dot{\hat{F}}_t^*(\theta^\dagger)(\hat{\theta} - \theta) \\
&= o_p\left(m^{-1/2}\right)
\end{aligned}$$

arguing as Robinson and Hidalgo (1997) with $\hat{\theta} - \theta = O_p(m^{-1/2})$, where m is the bandwidth parameter satisfying Assumption A4. Then, we have as $(N, T)_j \to \infty$,

$$\hat{F}_t^*(\hat{\theta}) - F_t^*(\theta) H_{F^*} = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right) + o_p(m^{-1/2})$$

because

$$\frac{1}{\min\left\{\sqrt{N},T\right\}} \approx \frac{1}{\sqrt{N}} + \frac{1}{T}$$

as $(N,T)_j \to \infty$. Furthermore,

$$\sqrt{N}\left(\hat{F}_t^*(\hat{\theta}) - F_t^*(\theta)H_{F^*}\right) = O_p(1) + O_p\left(\frac{\sqrt{N}}{T}\right) + o_p\left(\frac{\sqrt{N}}{\sqrt{m}}\right)$$

where the $O_p(1)$ term is the asymptotic normality result as shown below and the estimation error removal requires that $Nm^{-1} \to 0$ as $(N,T)_j \to \infty$.

Asymptotic normality is established, following similar steps to those used by Bai and Ng (2004), writing

$$\sqrt{N}\left(\hat{F}_t^*(\theta) - F_t^*(\theta)H_{F^*}\right) = \left(\frac{\Lambda^{*\prime}\Lambda^*}{N}\right)^{-1}\frac{1}{\sqrt{N}}\Lambda^{*\prime}\epsilon_t(\theta) + O_p\left(\frac{\sqrt{N}}{T}\right) + o_p\left(\frac{\sqrt{N}}{\sqrt{m}}\right)$$

leading to

$$\sqrt{N}\left(\hat{F}_t^*(\theta) - F_t^*(\theta)H_{F^*}\right) \to_d \mathcal{N}\left(0, \Sigma_{\Lambda^*}^{-1}\Gamma_t(\theta)\Sigma_{\Lambda^*}^{-1}\right)$$

for a fixed t if $Nm^{-1} \to 0$ as $(N,T)_j \to \infty$.$\square$

References

Abadir, K. M., Distaso, W., and Giraitis, L. (2007). Nonstationarity-extended local whittle estimation. *Journal of Econometrics*, 141(2):1353–1384.

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80(23):1806–1813.

Alonso, A. M., García-Martos, C., Rodríguez, J., and Jesús Sánchez, M. (2011). Seasonal dynamic factor analysis and bootstrap inference: application to electricity market forecasting. *Technometrics*, 53(2):137–151.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, pages 135–171.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration. *Econometrica*, pages 1127–1177.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends (R) in Econometrics*, 3(2):89–163.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2):409–424.

Breitung, J. and Eickmeier, S. (2016). Analyzing international business and financial cycles using multi-level factor models: A comparison of alternative approaches. *Dynamic Factor Models (Advances in Econometrics, Volume 35) Emerald Group Publishing Limited*, 35:177–214.

Breitung, J. and Pigorsch, U. (2013). A canonical correlation approach for selecting the number of dynamic factors. *Oxford Bulletin of Economics and Statistics*, 75(1):23–36.

Choi, I., Kim, D., Kim, Y. J., and Kwark, N.-S. (2016). A multilevel factor model: Identification, asymptotic theory and applications. Technical report.

Davidson, J. and Hashimzade, N. (2009). Type I and type II fractional brownian motions: A reconsideration. *Computational Statistics & Data Analysis*, 53(6):2089–2106.

Diebold, F. X., Li, C., and Yue, V. Z. (2008). Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach. *Journal of Econometrics*, 146(2):351–363.

Dordonnat, V., Koopman, S. J., and Ooms, M. (2012). Dynamic factors in periodic time-varying regressions with an application to hourly electricity load modelling. *Computational Statistics & Data Analysis*, 56(11):3134–3152.

Ergemen, Y. E. (2016). System estimation of panel data models under long-range dependence. *CREATES Research Paper*, 02.

Ergemen, Y. E., Haldrup, N., and Rodríguez-Caballero, C. V. (2015). Common long range dependence in a panel of hourly nord pool electricity prices and loads. *CREATES Research Paper*, 58.

Ergemen, Y. E. and Velasco, C. (2017). Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *Journal of Econometrics*, 196(2):248–258.

Gil-Alaña, L. A. and Robinson, P. M. (1997). Testing of unit root and other nonstationary hypotheses in macroeconomic time series. *Journal of Econometrics*, 80(2):241–268.

Haldrup, N. and Nielsen, M. Ø. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, 135(1):349–376.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Hallin, M. and Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, 163(1):29–41.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics*, 28(3):397–409.

Koopman, S. J., Ooms, M., and Carnero, M. A. (2007). Periodic seasonal reg-arfima–garch models for daily electricity spot prices. *Journal of the American Statistical Association*, 102(477):16–27.

Michelacci, C. and Zaffaroni, P. (2000). (Fractional) beta convergence. *Journal of Monetary Economics*, 45(1):129–153.

Moench, E., Ng, S., and Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, 95(5):1811–1817.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Pesaran, M. H. and Chudik, A. (2014). Aggregation in large dynamic panels. *Journal of Econometrics*, 178:273–285.

Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., and Brace, C. (1997). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting*, 13(2):161–174.

Raviv, E., Bouwman, K. E., and van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, 50:227 – 239.

Robinson, P. and Hidalgo, J. (1997). Time series regression with long-range dependence. *The Annals of Statistics*, 25(1):77–104.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, 26(02):501–540.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Wang, P. (2010). Large dimensional factor models with a multi-level factor structure: Identification, estimation, and inference. *Working paper. Hong Kong University of Science and Technology*.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*, volume 403. John Wiley & Sons.